

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
INSTITUTO DE MATEMÁTICA
DEPARTAMENTO DE ESTATÍSTICA

Time Series Studio: Uma interface gráfica para análise de séries temporais utilizando R e Shiny

Autor: *Pietro Tiaraju Giavarina dos Santos*
Orientador: *Prof. Dr. Eduardo de Oliveira Horta*

Porto Alegre, 30 de novembro de 2016.

Universidade Federal do Rio Grande do Sul

Instituto de Matemática e Estatística

Departamento de Estatística

Time Series Studio: Uma interface gráfica para análise de séries temporais utilizando R e Shiny

Autor: *Pietro Tiaraju Giavarina dos Santos*

Trabalho de Conclusão de Curso
apresentado para obtenção
do grau de Bacharel em Estatística.

Banca Examinadora:
Professor Dr. Eduardo de Oliveira Horta (IME - UFRGS)
Professor Dr. Cleber Bisognin (IME - UFRGS)
Flaviane Peccin Brevi (DSS - Souza Cruz)

Porto Alegre, 30 de novembro de 2016.

Dedico este trabalho ao mundo. Que cada uma das mais de 7 bilhões de pessoas existentes neste planeta possam de alguma forma se beneficiar de *Time Series Studio*.

Agradeço ao professor Eduardo Horta pelo apoio na realização e orientação desse trabalho.

Agradeço aos meus pais por tudo que fizeram por mim.

Agradeço àqueles amigos de verdade que ajudaram a construir a trajetória de minha vida até então, seja acadêmica ou pessoal.

E por que não agradecer também àqueles que de certa forma desagradei? É através das dificuldades, dos erros e das superações que aprendemos a ser uma pessoa melhor.

Agradeço também à equipe da Souza Cruz, em meu período de estágio, que me ajudou na elaboração do tema desse trabalho.

Também agradeço a mim mesmo pela persistência e dedicação ao longo dessa jornada.

E, dada a existência, finalizo agradecendo às forças que regem o Universo que fazem com que a vida seja tão bela e imprevisível.

Resumo

O presente trabalho irá mostrar as características teórico-práticas da interface gráfica *Time Series Studio*, desenvolvida pelo autor. O trabalho trouxe ao mundo uma ferramenta capaz de realizar uma modelagem completa em séries temporais. Dentre as vantagens de uso da interface gráfica, destacam-se: a grande economia de tempo em realizar uma modelagem, pois em menos de 5 minutos é possível obter resultados de previsão a partir dos dados temporais fornecidos à interface, e; a facilidade de uso, pois a única condição para a utilização da interface gráfica é uma conexão com a Internet. Dessa forma, não é necessária instalação prévia da interface no computador e também não requer do usuário conhecimentos de programação ou até mesmo sobre séries temporais. Sobre os recursos de modelagem, destacam-se o fornecimento de métodos automáticos para seleção de modelos, simulação de dados para uso educacional e previsão baseada no modelo selecionado, com a possibilidade de *download* de relatório-resumo com os principais resultados obtidos na modelagem.

A interface está disponível online no seguinte link:

<http://pietrotiaraju.shinyapps.io/Time_Series_Studio>

Palavras-chave: Interface Gráfica. Séries Temporais. R. Shiny.

Abstract

The present study will show the theoretical-practical characteristics of the Graphical User Interface (GUI), developed by the author. This study brought to the world a tool able to perform a complete modelling in time series. Among the advantages of using the GUI, the main highlights are: the big time saving in order to perform a modelling. It is possible to obtain forecasting results in less than 5 minutes based in data provided to the GUI, and; the ease of use, because the single condition to access the GUI is an Internet connection. In this way, it isn't necessary a previous instalation of the GUI in the computer and also it doesn't require programming knowledge and even about time series! About the modelling features, the main highlights are the automatic methods of model selection, data simulation for educational purposes and forecasting based in the selected model, with the possibility to download a summary-report which comprises the main results obtained in the modelling.

The GUI is available in the following link:

<http://pietrotiaraju.shinyapps.io/Time_Series_Studio>

Keywords: Graphical User Interface. Time Series. R. Shiny.

Sumário

1	Introdução	9
1.1	Contextualização	9
2	Elementos da Teoria de Séries Temporais	13
2.1	Definições Iniciais	13
2.2	Modelos de Séries Temporais	15
2.2.1	Modelo ARIMA	15
2.2.1.1	Modelo SARIMA	16
2.2.2	Modelo GARCH	17
2.2.3	Modelo ETS	17
2.2.4	Modelo Senoidal	20
2.3	Funções de Autocorrelação	20
2.4	Transformação de Dados	21
2.4.1	Testes Estatísticos	22
2.4.1.1	Teste de Phillips-Perron	22
2.4.1.2	Teste de Dickey-Fuller Aumentado	24
2.4.1.3	Teste de Box-Pierce	24
2.4.1.4	Teste de Shapiro-Wilk	24
2.5	Algoritmos de Seleção	25
2.5.1	Modelo SARIMA	25
2.5.2	Modelo GARCH	26
2.5.3	Modelo ETS	26
3	Guia do Usuário	28
3.1	<i>Import Data</i> (Importação de Dados)	28
3.2	<i>Simulate Data</i> (Simulação de Dados)	29
3.3	<i>Data Management</i> (Gestão de Dados)	29
3.4	<i>Analysis Tools</i> (Ferramentas de Análise)	30
3.5	<i>Model Adjustment</i> (Ajuste do Modelo)	32
3.6	<i>Residual Analysis</i> (Análise de Resíduos)	33
3.7	<i>Model Forecast</i> (Previsão do Modelo)	33
3.8	<i>Plot Options</i> (Opções Gráficas)	34
3.9	<i>Generate Report</i> (Geração de Relatório)	34

4	Considerações Finais	35
5	Anexos	36
6	Referências Bibliográficas	77

Capítulo 1

Introdução

A partir da constante evolução dos recursos computacionais voltados para a Estatística, surge também a oportunidade do desenvolvimento de ferramentas que tenham como objetivo auxiliar e facilitar tarefas com foco nesta área do conhecimento. Com a rarefeita quantidade de plataformas visuais disponíveis na área de séries temporais, o desenvolvimento de interfaces gráfica nessa área tornou-se um tema altamente relevante. Atrelado a esse crescimento de recursos computacionais, também apareceram opções interessantes para que o desenvolvedor possa exercer sua capacidade criativa. Baseado nessa iniciativa, o Shiny, um pacote estatístico disponível para a linguagem de programação R, fez com que se possa unir a motivação pela novidade com a possibilidade de se construir uma ferramenta útil ao estatístico, bem como a usuários oriundos de áreas afins.

1.1 Contextualização

A gênese da linguagem R é datada do início dos anos 90. A linguagem R é uma implementação da linguagem de programação S-Plus combinada com a semântica léxica de escopo inspirada em Scheme. Existem algumas diferenças importantes, mas boa parte do código inicial escrito em linguagem S-Plus foi inalterado. O R foi criado por Ross Ihaka e Robert Gentleman na Universidade de Auckland, Nova Zelândia, e atualmente é desenvolvido pela R Development Core Team.

O R e suas bibliotecas implementaram uma grande variedade de técnicas gráficas e estatísticas, incluindo modelagem linear e não-linear, testes estatísticos, análise de séries temporais, classificação, agrupamento, etc. O R é altamente

ampliável através de funções e extensões, além de sua comunidade ser notável pela ativa contribuição com o desenvolvimento de novos pacotes estatísticos. Por ser gratuito, tornou-se uma opção viável para ser o utilitário principal de diversos estatísticos, embora requeira conhecimentos de programação. (IHAKA, 1998 e MORANDAT, 2012).

Nos últimos anos, esse crescente aumento da popularidade do R fez com que se fossem desenvolvidas IDE's¹ para o mesmo. O objetivo da IDE é tornar a arte da programação uma experiência ainda mais satisfatória, facilitando o programador ao apresentar um visual mais robusto do que a interface gráfica original do R, limitada em recursos. O RStudio, lançado em 2011, é uma dessas IDE's existentes. (dos SANTOS et al., 2016).

Desenvolvido pelos mesmos criadores do RStudio, uma IDE que, nos últimos anos, se tornou um expoente para programadores em R devido a sua praticidade e ampla gama de recursos, o pacote Shiny fornece uma extensa bateria de elementos gráficos que podem ser dispostos de acordo com a preferência e necessidade do programador, permitindo a criação de ferramentas que qualquer usuário possa utilizar, sem a necessidade de possuir conhecimentos de programação ou sequer ter algum suporte ao R instalado no computador. De fato, uma conexão com a Internet é o suficiente para se ter acesso à uma interface Shiny, bastando apenas ter conexão com servidor em que o programa foi disponibilizado. Essa facilidade decorre do fato de que, embora a programação em Shiny seja dada em R, a interface é gerada em HTML, compatível com qualquer navegador que possua alguns recursos mínimos, como HTML5, JavaScript e CSS.

No contexto exposto acima, o presente trabalho tem como objetivos:

- i) introduzir o software Time Series Studio, desenvolvido pelo autor paralelamente à redação do presente texto;
- ii) apresentar a metodologia empregada na construção dessa interface, bem como as técnicas estatísticas utilizadas para realização das etapas de modelagem;
- iii) ilustrar o funcionamento da referida ferramenta, esclarecendo seu funcio-

¹Acônimo para *Integrated Development Environment*. Em português, Ambiente de Desenvolvimento Integrado.

namento e integração com as demais funções pré-existentes no CRAN² ou desenvolvidas pelo autor.

²Acrônimo para *The Comprehensive R Archive Network*. Em português, Rede Compreensiva de Arquivos R. O servidor disponibiliza todos os pacotes estatísticos oficializados, desenvolvidos por diversos autores independentes.

Capítulo 2

Elementos da Teoria de Séries Temporais

2.1 Definições Iniciais

Uma série temporal é um conjunto de dados dispostos em relação ao tempo. Geralmente ela é definida em intervalos de tempos iguais e sucessivos. Exemplos notórios e conhecidos de séries temporais vão desde contagem de manchas solares, marés oceânicas até fechamentos diários de uma bolsa de valores. Séries temporais são amplamente utilizadas em estatística, econometria, matemática financeira, astronomia, medicina, política, entre inúmeras outras áreas. Sendo um conceito multidisciplinar, sua abrangência prática é bastante significativa, mostrando-se necessário e inevitável para praticamente qualquer processo temporal conhecido no planeta. Nesse sentido, o conceito de Análise de Séries Temporais compreende diversas técnicas e abordagens para a realização de um melhor ajuste de modelo para esses dados temporais, com o objetivo de se extrair informações significativas desses dados, além de outras características importantes.

A análise de dados orientados no tempo e previsão de valores futuros de uma série temporal são dos problemas mais significativos que ocorrem em diversos campos da ciência. Das diversas razões que podem ser apresentadas para tal, destacam-se, entre outras:

- i) Administração de Operações: organizações empresariais rotineiramente usam previsões de vendas de produtos ou demanda por serviços com o

objetivo de agendar a produção, controlar inventário, gestão de cadeia de fornecedores, determinar necessidades de equipe de trabalho e planejar capacidade. Previsões podem ser utilizadas também para determinar o conjunto de produtos ou serviços que serão oferecidos e a localização da produção desses mesmos produtos;

- ii) Marketing: prever é extremamente importante em várias decisões de marketing. A previsão de resposta de vendas à anúncios, novas promoções ou mudança em políticas de preço permitem empresas a avaliarem sua efetividade no ramo, determinar os objetivos e metas que desejam alcançar e fazer qualquer tipo de ajuste;
- ii) Finanças e Gestão de Risco: investidores em ativos financeiros estão interessados em previsão dos retornos de seus investimentos. Outras decisões de investimento podem ser feitos relativos à previsão de taxa de juros, opções e variações cambiais;
- ii) Economia: governos, instituições financeiras e organizações políticas necessitam de previsões para variáveis de alta importância econômica, como PIB (Produto Interno Bruto), crescimento populacional, desemprego, taxa de juros, inflação, crescimento empregatício, produção e consumo. Estas previsões são parte integral da orientação através de políticas monetárias e fiscais, além de planos orçamentários e decisões estratégicas feitas pelo governo;
- ii) Controle Industrial de Processo: previsão de valores futuros de características de um processo de produção podem ajudar a determinar quando a mesma pode ser alterada, ou também quando o processo pode ser encerrado ou intensificado. Feedbacks e antecipações de projetos de controle são amplamente utilizados em monitoramento e ajuste de processos industriais. Além disso, predições da saída de processos também são parte integral desses projetos;
- ii) Demografia: previsões populacionais por país ou regiões administrativas são feitas rotineiramente, frequentemente estratificando por variáveis como gênero, idade e etnia. Demógrafos também prevêm nascimentos, mortes e padrões migratórios. Os governos utilizam estas previsões para planejar políticas e ações sociais, como o aumento de gastos na saúde, previdência ou até mesmo projetos de combate à pobreza. Muitos negócios utilizam

previsões de populações por faixa etária para elaborar planos estratégicos sobre o desenvolvimento de novas linhas de produto ou que tipos de serviço serão oferecidos.

A notação matemática de uma série temporal é da forma $\mathbf{X} = a_n \{x_t\}_{t=1}^n$, onde \mathbf{X} representa cada observação e n o número de observações. Cada observação é espaçada igualmente por um período de tempo. (MONTGOMERY, 2015 e IMDADULLAH, 2014).

Nesse contexto, a interface gráfica Time Series Studio oferece a possibilidade de uma modelagem extensiva em séries temporais baseada em um conjunto de técnicas estatísticas apropriadas, focada principalmente na transformação de dados, testes estatísticos, modelagem automática e previsão.

2.2 Modelos de Séries Temporais

Dentre diversos modelos de séries temporais existentes, quatro famílias de modelos serão contempladas neste trabalho: ARIMA, GARCH, ETS e Modelo Senoidal. Os dois primeiros estão contemplados tanto em estimação dos parâmetros quanto simulação, enquanto o terceiro apenas para estimação dos parâmetros e o último apenas em simulação.

2.2.1 Modelo ARIMA

Em estatística e econometria, o modelo ARIMA¹ é um dos mais utilizados na modelagem e previsões de séries temporais. O modelo foi sistematizado por George Box e Gwilym Jenkins, em 1976. (BOX e JENKINS, 1976). A parte autorregressiva (AR) indica que a variável de interesse é regredida nos seus próprios valores à priori. A parte de médias móveis (MA) indica que os erros da regressão são na verdade uma combinação de termos de erro cujos valores ocorrem contemporaneamente em vários momentos anteriores. A parte integrada (I) indica que os dados foram substituídos pela diferença de ordem n de seus valores, onde será posteriormente definida na seção de Transformação de Dados. O objetivo de cada uma dessas características é fazer com que o modelo ajuste os dados da melhor maneira possível. Em geral, os modelos ARIMA não-sazonais são definidos da seguinte forma: ARIMA(p, d, q), onde $p \geq 0$, $d \geq 0$,

¹Acrônimo para *Autoregressive Integrated Moving Average*, ou em português, Autorregressivo Integrado de Médias Móveis.

$q \succeq 0$, p é o grau do polinômio autorregressivo, d é o número de diferenciações (referente à n -ésima diferença simples) e q o grau do polinômio de médias móveis.

Dessa forma, pode-se apresentar um modelo ARIMA($p', 1, q'$) da seguinte forma, onde $\{X_t\}_{t \in \mathbb{Z}}$:

$$\left(1 - \sum_{i=1}^{p'} \alpha_i L^i\right) X_t = \left(1 - \sum_{i=1}^{q'} \theta_i L^i\right) \epsilon_t$$

onde L é o operador de defasagem, α_i os parâmetros do polinômio autorregressivo do modelo, θ_i são os parâmetros do polinômio de médias móveis e $\{\epsilon_t\}_{t \in \mathbb{Z}}$ os termos de erro.

Assumindo agora que o polinômio $(1 - \sum_{i=1}^{p'} \alpha_i L^i)$ tem raiz unitária (um fator $(1 - L)$) de multiplicidade d . Isso pode ser re-escrito como:

$$\left(1 - \sum_{i=1}^{p'} \alpha_i L^i\right) X_t = \left(1 - \sum_{i=1}^{p'-d} \phi_i L^i\right) (1 - L)^d X_t.$$

Um processo ARIMA(p, d, q) expressa a propriedade da fatoração polinomial com $p = p' - d$ e $q = q'$, e é dado por:

$$\left(1 - \sum_{i=1}^p \phi_i L^i\right) (1 - L)^d X_t = \left(1 - \sum_{i=1}^q \phi_i L^i\right) \epsilon_t.$$

O processo ainda pode ser generalizado para:

$$\left(1 - \sum_{i=1}^p \phi_i L^i\right) (1 - L)^d X_t = \delta + \left(1 - \sum_{i=1}^q \phi_i L^i\right) \epsilon_t.$$

onde δ é o *drift* do processo.

2.2.1.1 Modelo SARIMA

No objetivo de estender a amplitude e as possibilidade da modelagem ARIMA, é possível acrescentar um componente sazonal nos modelos ARIMA. Os modelos da família SARIMA são definidos da seguinte forma: SARIMA(p, d, q) \times (P, D, Q) $_m$ onde m refere-se ao número de períodos em cada temporada, ou mais especificamente, o grau do polinômio. Também $P \succeq 0$, $D \succeq 0$, $Q \succeq 0$, onde P é o grau do polinômio autorregressivo sazonal, D é o número de diferenciações (referente à n -ésima diferença sazonal) e Q a ordem do polinômio de médias móveis sazonal. A equação do modelo é generalizada abaixo:

$$\left(1 - \sum_{i=1}^p \phi_i L^i\right) \left(1 - \sum_{i=1}^P \Phi_i L^i\right) (1-L)^d (1-L)^D X_t = \delta + \left(1 - \sum_{i=1}^q \phi_i L^i\right) \left(1 - \sum_{i=1}^Q \Theta_i L^i\right) \epsilon_t.$$

2.2.2 Modelo GARCH

Se um modelo ARIMA for assumido para a variância dos erros, então o novo modelo de séries temporais é chamado de modelo GARCH². GARCH é a condição para uma ou mais observações na série para a qual a variância do termo de erro corrente ou inovação é função dos tamanhos atuais dos termos de erro dos períodos de tempo anteriores: frequentemente a variância é relacionada aos quadrados das inovações anteriores. Seu principal uso é na área da econometria, principalmente para a modelagem de séries de retornos financeiros de ações. Neste caso, o modelo GARCH(p, q), onde p é a ordem do número de termos GARCH σ^2 e q a ordem dos termos ARCH ϵ^2 . A notação matemática de um processo GARCH(p, q) é definida como:

$$\sigma_t^2 = \omega + \alpha_1 \epsilon_{t-1}^2 + \dots + \alpha_q \epsilon_{t-q}^2 + \beta_1 \sigma_{t-1}^2 + \dots + \beta_p \sigma_{t-p}^2 = \omega + \sum_{i=1}^q \alpha_i \epsilon_{t-i}^2 + \sum_{i=1}^p \beta_i \sigma_{t-i}^2$$

onde $\{\epsilon_t\}_{t \in \mathbb{Z}}$ é ruído branco, $\{\alpha_1, \dots, \alpha_q\} > 0$ são os parâmetros associados aos quadrados do ruído branco, $\{\beta_1, \dots, \beta_p\} > 0$ os parâmetros associados à variância e $\omega > 0$ a constante associada ao modelo.

Geralmente, quando testa-se por heteroscedasticidade em modelos econométricos, o mais comumente utilizado é o Teste de White, que neste trabalho não será inicialmente abordado, ficando para uma futura atualização da interface gráfica, assim como outras variações dos modelos GARCH.

2.2.3 Modelo ETS

A linhagem de modelos ETS³ é menos comumente utilizada que as anteriores. Cada modelo consiste em uma equação que descreve os dados observados e algumas equações de transição que descrevem como os componentes não observados ou estados (nível, tendência e sazonalidade) se modificam ao longo do tempo.

²acrônimo para *Generalized Autoregressive Conditional Heteroskedasticity*. Em português, Heteroscedasticidade Condicional Autoregressiva Generalizada.

³Acronimo para *Exponential Smoothing*. Em português, Suavização Exponencial.

Seus métodos de ajuste são datados desde a década de 50, os quais foram inicialmente classificados por Pegels (1969), estendidos ao longo do tempo por Gardner (1985), Hyndman et al. (2002) e estendido novamente por Taylor (2003), onde é possível verificar as possibilidades de ajuste na Figura 2.1. (HYNDMAN e KHANDAKHAR, 2002).

ADDITIVE ERROR MODELS

Trend	Seasonal		
	N	A	M
N	$y_t = \ell_{t-1} + \varepsilon_t$ $\ell_t = \ell_{t-1} + \alpha\varepsilon_t$	$y_t = \ell_{t-1} + s_{t-m} + \varepsilon_t$ $\ell_t = \ell_{t-1} + \alpha\varepsilon_t$ $s_t = s_{t-m} + \gamma\varepsilon_t$	$y_t = \ell_{t-1}s_{t-m} + \varepsilon_t$ $\ell_t = \ell_{t-1} + \alpha\varepsilon_t/s_{t-m}$ $s_t = s_{t-m} + \gamma\varepsilon_t/\ell_{t-1}$
A	$y_t = \ell_{t-1} + b_{t-1} + \varepsilon_t$ $\ell_t = \ell_{t-1} + b_{t-1} + \alpha\varepsilon_t$ $b_t = b_{t-1} + \beta\varepsilon_t$	$y_t = \ell_{t-1} + b_{t-1} + s_{t-m} + \varepsilon_t$ $\ell_t = \ell_{t-1} + b_{t-1} + \alpha\varepsilon_t$ $b_t = b_{t-1} + \beta\varepsilon_t$ $s_t = s_{t-m} + \gamma\varepsilon_t$	$y_t = (\ell_{t-1} + b_{t-1})s_{t-m} + \varepsilon_t$ $\ell_t = \ell_{t-1} + b_{t-1} + \alpha\varepsilon_t/s_{t-m}$ $b_t = b_{t-1} + \beta\varepsilon_t/s_{t-m}$ $s_t = s_{t-m} + \gamma\varepsilon_t/(\ell_{t-1} + b_{t-1})$
A _d	$y_t = \ell_{t-1} + \phi b_{t-1} + \varepsilon_t$ $\ell_t = \ell_{t-1} + \phi b_{t-1} + \alpha\varepsilon_t$ $b_t = \phi b_{t-1} + \beta\varepsilon_t$	$y_t = \ell_{t-1} + \phi b_{t-1} + s_{t-m} + \varepsilon_t$ $\ell_t = \ell_{t-1} + \phi b_{t-1} + \alpha\varepsilon_t$ $b_t = \phi b_{t-1} + \beta\varepsilon_t$ $s_t = s_{t-m} + \gamma\varepsilon_t$	$y_t = (\ell_{t-1} + \phi b_{t-1})s_{t-m} + \varepsilon_t$ $\ell_t = \ell_{t-1} + \phi b_{t-1} + \alpha\varepsilon_t/s_{t-m}$ $b_t = \phi b_{t-1} + \beta\varepsilon_t/s_{t-m}$ $s_t = s_{t-m} + \gamma\varepsilon_t/(\ell_{t-1} + \phi b_{t-1})$
M	$y_t = \ell_{t-1}b_{t-1} + \varepsilon_t$ $\ell_t = \ell_{t-1}b_{t-1} + \alpha\varepsilon_t$ $b_t = b_{t-1} + \beta\varepsilon_t/\ell_{t-1}$	$y_t = \ell_{t-1}b_{t-1} + s_{t-m} + \varepsilon_t$ $\ell_t = \ell_{t-1}b_{t-1} + \alpha\varepsilon_t$ $b_t = b_{t-1} + \beta\varepsilon_t/\ell_{t-1}$ $s_t = s_{t-m} + \gamma\varepsilon_t$	$y_t = \ell_{t-1}b_{t-1}s_{t-m} + \varepsilon_t$ $\ell_t = \ell_{t-1}b_{t-1} + \alpha\varepsilon_t/s_{t-m}$ $b_t = b_{t-1} + \beta\varepsilon_t/(\ell_{t-1}s_{t-m})$ $s_t = s_{t-m} + \gamma\varepsilon_t/(\ell_{t-1}b_{t-1})$
M _d	$y_t = \ell_{t-1}b_{t-1}^\phi + \varepsilon_t$ $\ell_t = \ell_{t-1}b_{t-1}^\phi + \alpha\varepsilon_t$ $b_t = b_{t-1}^\phi + \beta\varepsilon_t/\ell_{t-1}$	$y_t = \ell_{t-1}b_{t-1}^\phi + s_{t-m} + \varepsilon_t$ $\ell_t = \ell_{t-1}b_{t-1}^\phi + \alpha\varepsilon_t$ $b_t = b_{t-1}^\phi + \beta\varepsilon_t/\ell_{t-1}$ $s_t = s_{t-m} + \gamma\varepsilon_t$	$y_t = \ell_{t-1}b_{t-1}^\phi s_{t-m} + \varepsilon_t$ $\ell_t = \ell_{t-1}b_{t-1}^\phi + \alpha\varepsilon_t/s_{t-m}$ $b_t = b_{t-1}^\phi + \beta\varepsilon_t/(s_{t-m}\ell_{t-1})$ $s_t = s_{t-m} + \gamma\varepsilon_t/(\ell_{t-1}b_{t-1}^\phi)$

MULTIPLICATIVE ERROR MODELS

Trend	Seasonal		
	N	A	M
N	$y_t = \ell_{t-1}(1 + \varepsilon_t)$ $\ell_t = \ell_{t-1}(1 + \alpha\varepsilon_t)$	$y_t = (\ell_{t-1} + s_{t-m})(1 + \varepsilon_t)$ $\ell_t = \ell_{t-1} + \alpha(\ell_{t-1} + s_{t-m})\varepsilon_t$ $s_t = s_{t-m} + \gamma(\ell_{t-1} + s_{t-m})\varepsilon_t$	$y_t = \ell_{t-1}s_{t-m}(1 + \varepsilon_t)$ $\ell_t = \ell_{t-1}(1 + \alpha\varepsilon_t)$ $s_t = s_{t-m}(1 + \gamma\varepsilon_t)$
A	$y_t = (\ell_{t-1} + b_{t-1})(1 + \varepsilon_t)$ $\ell_t = (\ell_{t-1} + b_{t-1})(1 + \alpha\varepsilon_t)$ $b_t = b_{t-1} + \beta(\ell_{t-1} + b_{t-1})\varepsilon_t$	$y_t = (\ell_{t-1} + b_{t-1} + s_{t-m})(1 + \varepsilon_t)$ $\ell_t = \ell_{t-1} + b_{t-1} + \alpha(\ell_{t-1} + b_{t-1} + s_{t-m})\varepsilon_t$ $b_t = b_{t-1} + \beta(\ell_{t-1} + b_{t-1} + s_{t-m})\varepsilon_t$ $s_t = s_{t-m} + \gamma(\ell_{t-1} + b_{t-1} + s_{t-m})\varepsilon_t$	$y_t = (\ell_{t-1} + b_{t-1})s_{t-m}(1 + \varepsilon_t)$ $\ell_t = (\ell_{t-1} + b_{t-1})(1 + \alpha\varepsilon_t)$ $b_t = b_{t-1} + \beta(\ell_{t-1} + b_{t-1})\varepsilon_t$ $s_t = s_{t-m}(1 + \gamma\varepsilon_t)$
A _d	$y_t = (\ell_{t-1} + \phi b_{t-1})(1 + \varepsilon_t)$ $\ell_t = (\ell_{t-1} + \phi b_{t-1})(1 + \alpha\varepsilon_t)$ $b_t = \phi b_{t-1} + \beta(\ell_{t-1} + \phi b_{t-1})\varepsilon_t$	$y_t = (\ell_{t-1} + \phi b_{t-1} + s_{t-m})(1 + \varepsilon_t)$ $\ell_t = \ell_{t-1} + \phi b_{t-1} + \alpha(\ell_{t-1} + \phi b_{t-1} + s_{t-m})\varepsilon_t$ $b_t = \phi b_{t-1} + \beta(\ell_{t-1} + \phi b_{t-1} + s_{t-m})\varepsilon_t$ $s_t = s_{t-m} + \gamma(\ell_{t-1} + \phi b_{t-1} + s_{t-m})\varepsilon_t$	$y_t = (\ell_{t-1} + \phi b_{t-1})s_{t-m}(1 + \varepsilon_t)$ $\ell_t = (\ell_{t-1} + \phi b_{t-1})(1 + \alpha\varepsilon_t)$ $b_t = \phi b_{t-1} + \beta(\ell_{t-1} + \phi b_{t-1})\varepsilon_t$ $s_t = s_{t-m}(1 + \gamma\varepsilon_t)$
M	$y_t = \ell_{t-1}b_{t-1}(1 + \varepsilon_t)$ $\ell_t = \ell_{t-1}b_{t-1}(1 + \alpha\varepsilon_t)$ $b_t = b_{t-1}(1 + \beta\varepsilon_t)$	$y_t = (\ell_{t-1}b_{t-1} + s_{t-m})(1 + \varepsilon_t)$ $\ell_t = \ell_{t-1}b_{t-1} + \alpha(\ell_{t-1}b_{t-1} + s_{t-m})\varepsilon_t$ $b_t = b_{t-1} + \beta(\ell_{t-1}b_{t-1} + s_{t-m})\varepsilon_t/\ell_{t-1}$ $s_t = s_{t-m} + \gamma(\ell_{t-1}b_{t-1} + s_{t-m})\varepsilon_t$	$y_t = \ell_{t-1}b_{t-1}s_{t-m}(1 + \varepsilon_t)$ $\ell_t = \ell_{t-1}b_{t-1}(1 + \alpha\varepsilon_t)$ $b_t = b_{t-1}(1 + \beta\varepsilon_t)$ $s_t = s_{t-m}(1 + \gamma\varepsilon_t)$
M _d	$y_t = \ell_{t-1}b_{t-1}^\phi(1 + \varepsilon_t)$ $\ell_t = \ell_{t-1}b_{t-1}^\phi(1 + \alpha\varepsilon_t)$ $b_t = b_{t-1}^\phi(1 + \beta\varepsilon_t)$	$y_t = (\ell_{t-1}b_{t-1}^\phi + s_{t-m})(1 + \varepsilon_t)$ $\ell_t = \ell_{t-1}b_{t-1}^\phi + \alpha(\ell_{t-1}b_{t-1}^\phi + s_{t-m})\varepsilon_t$ $b_t = b_{t-1}^\phi + \beta(\ell_{t-1}b_{t-1}^\phi + s_{t-m})\varepsilon_t/\ell_{t-1}$ $s_t = s_{t-m} + \gamma(\ell_{t-1}b_{t-1}^\phi + s_{t-m})\varepsilon_t$	$y_t = \ell_{t-1}b_{t-1}^\phi s_{t-m}(1 + \varepsilon_t)$ $\ell_t = \ell_{t-1}b_{t-1}^\phi(1 + \alpha\varepsilon_t)$ $b_t = b_{t-1}^\phi(1 + \beta\varepsilon_t)$ $s_t = s_{t-m}(1 + \gamma\varepsilon_t)$

Figura 2.1: Possibilidades de modelos ETS

No artigo *Automatic Time Series Forecasting: The forecast package of R* de Hyndman et al. (2002) é possível avaliar um panorama mais avançado sobre

modelagem exponencial e a linhagem de modelos ETS.

2.2.4 Modelo Senoidal

O modelo senoidal, utilizado em diversos ramos como estatística, processamento de sinais e análise de séries temporais para aproximar uma sequência Y_i satisfaz a seguinte equação, com $i = \{1, \dots, n\}$ correspondente à ordem do modelo:

$$Y_i = C + \alpha \sin(\omega T_i + \phi) + E_i$$

onde C é uma constante que define o nível médio, α é a amplitude da onda senoidal, ω é a frequência, T_i é a variável temporal, ϕ a fase e E_i a sequência dos erros devido à aproximação da sequência Y_i pelo modelo. O modelo senoidal também pode ser ajustado utilizando mínimos quadrados não-lineares. Para se obter um bom ajuste, rotinas para mínimos quadrados não-lineares podem requerer bons valores iniciais para a constante, amplitude e frequência. (INFORMATION TECHNOLOGY LABORATORY, 2016).

2.3 Funções de Autocorrelação

Dentre as variadas visões disponíveis em análise de séries temporais, destacam-se as funções de autocorrelação e autocorrelação parcial. Utilizado notoriamente em modelagem no método Box-Jenkins para processos ARIMA.

De modo geral, a função de autocorrelação amostral e a função de autocorrelação parcial amostral são comparadas com o seu esperado comportamento teórico. Especificamente, para um modelo ARIMA(1, 0, 0), ou simplificadaamente AR(1), por exemplo, a função de autocorrelação deveria ter um comportamento exponencialmente decrescente. Entretanto, para processos AR de ordem superior as funções de autocorrelação aparecem frequentemente como uma mistura de decaimento exponencial e componentes senoidais amortecidos.

Também para processos AR de ordem superior, a função de autocorrelação necessita ser complementada com a função de autocorrelação parcial. A função de autocorrelação parcial para o modelo AR(p) será zero até *lags* maiores que $p + 1$, então é necessário examinar a função de autocorrelação parcial para verificar se há evidência de afastamento do zero nos *lags* correspondentes. Isso é usualmente determinado através do desenho de intervalos de confiança a 95% na função de autocorrelação.

Formato	Modelo indiciado
Exponencial, decaindo a zero	Modelo autorregressivo: $ARIMA(p, 0, 0)$ ou $AR(p)$
Alternando entre positivo e negativo, decaindo a zero	Modelo autorregressivo: $ARIMA(p, 0, 0)$ ou $AR(p)$
Picos isolados, com o restante sendo essencialmente zero	Modelo de médias móveis: $ARIMA(0, 0, q)$ ou $MA(q)$
Decaimento a partir de alguns <i>lags</i>	Modelo autorregressivo e médias móveis: $ARIMA(p, 0, q)$ ou $ARMA(p, q)$
Tudo zero ou próximo de zero	Ruído branco: $ARIMA(0, 0, 0)$
Picos altos em intervalos fixos	Modelo autorregressivo sazonal: $SARIMA(p, 0, q)(P, 0, Q)_m$
Sem decaimento a zero	Série não-estacionária: $ARIMA(p, d, q)$ ou $SARIMA(p, d, q)(P, D, Q)_m$ com $d \geq 1$ e $D \geq 1$

Tabela 2.1: Tabela-resumo com diagnóstico de funções de autocorrelação

Já para processos do tipo $ARIMA(0, 0, q)$, ou $MA(q)$, a função de autocorrelação será zero a partir do lag $q + 1$, então deve ser examinado onde a função de autocorrelação se torna essencialmente zero. O procedimento de traçar intervalos de confiança também deve ser novamente utilizado para o mesmo fim. Porém, a função de autocorrelação parcial não é útil para identificar processos do tipo MA . Na Tabela 2.1, segue alguns possíveis diagnósticos das aparências possíveis das funções de autocorrelação.

2.4 Transformação de Dados

Transformações de dados em série temporais podem ser úteis no intuito de remover a heteroscedasticidade dos dados, que pode ser detectado de maneira visual e também analisando os respectivos correlogramas. É oferecida a possi-

bilidade de aplicar uma transformação do tipo $f(x)$, onde x é a série temporal inicialmente definida. Dentre transformações utilizadas para esse fim são mais usuais $f(x) = \ln(x)$ e $f(x) = \sqrt{x}$. Diferenças simples e sazonais nos dados também podem ser aplicadas de forma que possa ser visualizada a remoção da tendência da série temporal, caso houver. A tendência da série temporal refere-se à mudança de distribuição dos dados ao longo do tempo, ou seja, quando os dados não estão centrados ou variando em torno de uma mesma média. Sendo assim, a série da primeira diferença simples é definida por $\delta_t = x_t - x_{t-1}$, onde $t = \{0, 1, 2, \dots, (n-1)\}$ e x a observação correspondente da série temporal original. A nova série temporal terá $n-1$ observações. Sucessivamente, séries de diferenças maior que 1 seguem a mesma aplicação. Uma série temporal da segunda diferença simples é definida por: $\delta_t^2 = \delta_t - \delta_{t-1}$, onde $t = \{0, 1, 2, \dots, (n-1)\}$ e δ a observação corresponde da série da primeira diferença simples. Diferenças sazonais são conceitualmente similares à diferenças simples, mas se contrapõem no sentido que, ao invés de se tomar a diferença da próxima observação com a observação anterior, a diferença tomada é baseada no domínio da frequência, ou seja, dependendo do período em que os dados são observados, se tomará a diferença apropriada com um *lag* específico. Sendo assim, definimos diferença sazonal como: $s_t = x_t - x_{t-p}$, onde $t = \{0, 1, 2, \dots, (n-1)\}$, x a observação correspondente da série temporal original e p o período sazonal (sazonalidade) da série temporal.

2.4.1 Testes Estatísticos

A possibilidade de realização de testes estatísticos apropriados para cada etapa da modelagem é também de bastante valia para uma análise completa. Dessa forma, é possível comparar o resultado de um teste estatístico utilizando a série temporal original com o resultado do teste numa série temporal transformada e assim, com base nisso, tomar alguma atitude a respeito sobre a transformação dos dados. Dentre os testes estatísticos contemplados pré-modelagem, são oferecidos: Teste de Phillips-Perron, Teste de Dickey-Fuller Aumentado e Teste de Box-Pierce.

2.4.1.1 Teste de Phillips-Perron

Este teste é utilizado em análise de séries temporais com o objetivo de testar a hipótese nula de que uma série temporal é integrada de ordem 1. É uma

generalização do Teste de Dickey-Fuller para os casos em que os erros $\{\epsilon_t\}_{t \in Z}$ são correlacionados e, possivelmente, heteroscedásticos. Considerando a seguinte regressão:

$$\Delta y_t = \beta_1 + \beta_2 t + \delta y_{t-1} + \sum_{i=1}^m \alpha_i \Delta y_{t-i} + \epsilon_t \quad (2.1)$$

onde β_1 é o intercepto, também denominado de drift da série; β_2 é o coeficiente de tendência; δ é coeficiente de presença de raiz unitária e m o número de defasagens tomada da série. Neste caso, a estatística de teste Z é calculada por:

$$Z = n\hat{\delta}_n - \frac{n^2 + \hat{\sigma}^2}{2s_n^2} \left(\hat{\lambda}_n^2 - \hat{\gamma}_{0,n} \right),$$

onde

$$\hat{\gamma}_{j,n} = \frac{1}{n} \sum_{i=1+j}^n r_i r_{i-j},$$

$$\hat{\lambda}_n^2 = \hat{\gamma}_{0,n} + 2 \sum_{j=1}^q \left(1 - \frac{j}{q+1} \right) \hat{\gamma}_{j,n}$$

e

$$s_n^2 = \frac{1}{(n-k)} \sum_{i=1}^n r_i^2.$$

Em que r_i representa o resíduo em y_i utilizando estimadores de mínimos quadrados, k é o número de covariáveis na regressão e, q é o número de defasagens utilizadas para calcular $\hat{\lambda}_n^2$. Caso o processo não seja correlacionado temos covariâncias nulas e neste caso, $\hat{\lambda}_n^2 = \hat{\gamma}_{0,n}$. Se o processo não for heteroscedástico temos que $ep(\delta) = \frac{1}{n}$ e então Z é dada por:

$$Z = n\hat{\delta} = \frac{\hat{\delta}}{ep(\hat{\delta})},$$

onde $\hat{\delta}$ é um estimador δ e $ep(\hat{\delta})$ um estimador para o desvio padrão de $\hat{\delta}$. Dessa forma, Z é a estatística de Dickey-Fuller e portanto, tem a mesma distribuição da estatística do Teste de Dickey-Fuller Aumentado, calculada por Dickey-Fuller através de simulação de Monte Carlo.

2.4.1.2 Teste de Dickey-Fuller Aumentado

O Teste de Dickey-Fuller Aumentado é conhecido na literatura como Teste ADF (Augmented Dickey-Fuller) e requer o estudo sobre a regressão apresentada na Equação 2.1, bem como suas sub-definições. Neste caso, a hipótese nula é dada por: $H_0 : \delta = 0$. Sendo assim, é realizada uma regressão de Δy_t em y_{t-1} , $\Delta y_{t-1}, \dots, \Delta y_{t+p-1}$ e calculada a estatística T dada por:

$$T = \frac{\hat{\delta}}{ep(\hat{\delta})},$$

onde $\hat{\delta}$ e $ep(\hat{\delta})$ estão previamente definidos na subsubseção anterior. Os valores críticos da estatística T foram tabelados por Dickey-Fuller através de simulação de Monte Carlo e variam nos casos de presença somente de intercepto, presença somente de tendência e presença de ambos.

2.4.1.3 Teste de Box-Pierce

Box e Pierce sugeriram um teste para as auto-correlações seriais, o qual indica se estes valores estão muito altos. Inicialmente, é calculada uma estimativa para as auto-correlações por:

$$\hat{r}_k = \frac{\sum_{t=k+1}^n \hat{a}_t \hat{a}_{t-k}}{\sum_{t=1}^n \hat{a}_t^2},$$

Se o modelo for apropriado, a estatística de teste

$$Q(k) = n \sum_{j=1}^K \hat{r}_j^2$$

terá aproximadamente uma distribuição χ^2 com K graus de liberdade, onde K é o número de defasagens tomada na função de autocorrelação. Portanto, rejeitamos a hipótese nula de que os erros são ruído branco se $Q > \chi_{1-\alpha, k}^2$ com um nível de significância α .

2.4.1.4 Teste de Shapiro-Wilk

O Teste de Shapiro-Wilk, proposto em 1965, é baseado na estatística W dada por:

$$W = \frac{b^2}{\sum_{i=1}^n (x_{(i)} - \bar{x})^2}$$

em que $x_{(i)}$ são os valores da amostra ordenados. A constante b é determinada da seguinte forma:

$$b = \begin{cases} \sum_{i=1}^{n/2} a_{n-i+1} (x_{(n-i+1)} - x_{(i)}) & \text{se } n \text{ é par,} \\ \sum_{i=1}^{(n+1)/2} a_{n-i+1} (x_{(n-i+1)} - x_{(i)}) & \text{se } n \text{ é ímpar} \end{cases}$$

onde a_{n-i+1} são constantes geradas pelas médias, variâncias e covariâncias das estatísticas de ordem de uma amostra de tamanho n de uma distribuição normal.

Assim, para a realização do teste, deve-se considerar a hipótese nula de que a amostra provém de uma distribuição normal contra uma hipótese alternativa de que a mesma não vem de distribuição normal. Conforme o nível de significância α definido, rejeita-se ou não a hipótese nula dependendo do valor de W_{calc} , que é a estatística W calculada para a amostra de teste e W_{tab} , o valor de W tabelado que é variável com o tamanho amostral n e o nível de significância α .

2.5 Algoritmos de Seleção

A determinação do modelo que melhor se ajusta aos dados pode ser definido de diversas formas e critérios, como será visto nessa seção.

2.5.1 Modelo SARIMA

A família de modelos SARIMA é predita pela interface gráfica através de um algoritmo desenvolvido por HYNDMAN et al., 2008, que combina testes de raiz unitárias, minimização do AIC⁴ e mínimos quadrados dos resíduos com o objetivo de se encontrar o melhor modelo.

O algoritmo procede da seguinte forma:

- i) O número de diferenças d é determinado usando em sucessivas repetições o Teste de Kwiatkowski-Phillips-Schmidt-Shin (KPSS)⁵;

⁴ Abreviação de *Akaike Information Criterion*, ou em português, Critério de Informação de Akaike. O critério é uma medida de qualidade relativa ao ajuste de um modelo, onde leva-se em conta a significância das estimativas e a quantidade de parâmetros. (HU, 2007).

⁵ O Teste KPSS é utilizado para testar a hipótese nula de que uma série temporal observada

- ii) Os valores de p e q são escolhidos através da minimização do AIC após a diferenciação dos dados d vezes. Ao invés de considerar cada combinação de p e q , o algoritmo utiliza um procedimento de pesquisa stepwise para percorrer o espaço do modelo, repetindo sucessivas vezes até verificar o menor AIC.

2.5.2 Modelo GARCH

O melhor modelo GARCH é escolhido através de metodologia desenvolvida pelo autor da interface gráfica, devido a ausência de algoritmos de seleção disponíveis no CRAN e/ou em repositórios similares. O algoritmo procede a escolha do modelo da seguinte forma:

- i) São consideradas 20 combinações $\text{GARCH}(p, q)$, onde $p = \{1, 2, 3, 4\}$ e $q = \{0, 1, 2, 3, 4\}$;
- ii) A partir dessas 20 combinações, o algoritmo testa para cada valor de p e q , o seu respectivo AIC e o número de valores ajustados significativos a nível de 95% de confiança. Desta forma, seleciona o modelo que tenha o menor AIC e que possua todas suas estimativas significativas.

2.5.3 Modelo ETS

O melhor ajuste para a família de modelos ETS é descrito da seguinte forma, por HYNDMAN, 2008:

- i) Para cada série, é aplicado todos os modelos, otimizando os parâmetros (ambos parâmetros de suavização e variável de estado inicial) do modelo em cada caso;
- ii) O algoritmo seleciona o melhor modelo de acordo com o AIC;
- iii) São produzidas previsões pontuais utilizando o melhor modelo (com os parâmetros otimizados);

é estacionária em volta de uma tendência determinística contra uma hipótese alternativa de raiz unitária.

- iv) Obtêm-se então intervalos de predição para o melhor modelo tanto utilizando resultados analíticos, ou por simulação de caminhos amostrais futuros para $\{y_{n+1}, \dots, y_{n+h}\}$ e encontrando os percentis $(\alpha/2)100\%$ e $(1 - \alpha/2)100\%$ dos dados simulados para cada janela de predição. Se a simulação é utilizada, os caminhos amostrais podem ser gerados utilizando a distribuição normal para erros (*bootstrap* paramétrico) ou utilizando os erros reamostrados (*bootstrap* ordinário).

Capítulo 3

Guia do Usuário

A interface gráfica consiste em um painel lateral e um painel principal onde são realizadas todas as operações, sejam de entrada ou saída de dados. No menu lateral, há uma série de opções sequenciais a serem selecionadas. Elas serão detalhadas operacionalmente neste capítulo.

Todas as Figuras dessa seção estão disponíveis no Anexo deste trabalho.

3.1 *Import Data* (Importação de Dados)

A tela inicial do programa dispõe da seção *Import Data*, a qual consiste em três sub-menus, sendo uma de saída de dados, tendo elas funcionalidades distintas. A primeira sub-estrutura de entrada, *Data Selector*, é onde o usuário poderá realizar a importação de uma série temporal pré-existente, por meio de um arquivo em formato XLS ou XLSX (Planilha do Microsoft Excel). No segundo sub-menu, a caixa correspondente à *Sheet Selector*, é oferecida a possibilidade do usuário selecionar a planilha correspondente ao conjunto de dados desejados, caso haja mais de uma no arquivo selecionado. Por último, na *Viewing Pane*, é possível visualizar uma prévia dos dados. Na Figura 5.1 é possível observar a tela inicial do programa.

Será utilizado para mostrar a ferramenta a série temporal da produção mensal de eletricidade na Austrália (em milhões de kWh), no período compreendido entre janeiro de 1956 e agosto de 1995¹. A Figura 5.2 mostra como é feita a

¹retirado da página DataMarket. Disponível em: <https://datamarket.com/data/set/2210/monthly-electricity-production-in-australia-million-kilowatt-hours-jan-1956-aug-1995>.

importação dos dados temporais na interface.

Por fim, ainda na mesma tela, a Figura 5.3 nos mostra o menu após a importação dos dados, com a *Sheet Selector* e o *Viewing Pane* ativado mostrando a planilha ativa e os dados já carregados na interface.

3.2 *Simulate Data* (Simulação de Dados)

Alternativamente à importação de dados, é possível através dessa opção realizar a simulação de dados para fins didáticos caso o usuário ache conveniente. Esse painel consiste em três sub-menus. Os dados podem ser gerados a partir de três famílias distintas de modelos: ARIMA, GARCH e Modelo Senoidal, onde cada um deles têm seus parâmetros específicos para a geração apropriada desses dados. Também é oferecida a possibilidade de escolher o número de observações a serem simuladas. Haverá uma tela abaixo para a visualização da série temporal gerada. A geração de dados para a família de modelos ARIMA é efetuada utilizando a função *arima.sim*, disponível no pacote *stats*. Já a geração de dados para a família de modelos GARCH é feita utilizando a função *garchSim*, do pacote *fGarch*. Essa função também contempla simulação de modelo misto ARIMA-GARCH, porém inicialmente não será abordada na interface. Por último, a simulação de Modelo Senoidal cujo código-fonte para a simulação deste modelo foi programado pelo autor. Nas Figuras 5.4, 5.5 e 5.6 são apresentadas simulações para os três tipos de modelos disponíveis.

3.3 *Data Management* (Gestão de Dados)

Na seção de *Data Management*, é possível definir algumas propriedades básicas da série temporal. Até então, o que se havia para a interface gráfica era um simples conjunto de dados sem associação temporal específica. Logo, é primordial para o prosseguimento da análise a utilização desse menu. Um dos sub-menus, *Select Data*, é habilitado apenas quando o usuário realizou a importação de um conjunto de dados pré-existente. Como o arquivo poderá conter mais de um conjunto de dados, é oferecida a possibilidade de definição da variável desejada a ser modelada, facilitando também com que o usuário possa aglutinar diversos conjuntos de dados em apenas uma única planilha. Também é disponibilizado um menu com três abas para a definição temporal da série de dados: *Frequency*

(Frequência), *Start Year* (Ano de Início) e *Start Period* (Período de Início). A primeira aba diz respeito à frequência em que os dados foram extraídos. Por padrão, o seletor vem definido como *Monthly* (mensal), ou seja, considerando uma série temporal com dados mensais (frequência = 12, ou 12 vezes ao ano). Outras opções são fornecidas como: *Daily* (diário), *Quarterly* (trimestral), *Semiannual* (semestral), *Annual* (anual) e *User defined* (definido pelo usuário). Na última opção é possível manualmente definir um período diferente dos oferecidos como padrão. A segunda aba, *Start Year*, corresponde ao ano de início da série temporal e a terceira aba corresponde ao período de início da série temporal. Ou seja, se os dados iniciaram no 3º mês de uma série com dados mensais, deve ser escolhido o valor 3. Uma *Viewing Pane* será disponibilizada para o usuário visualizar os eixos, agora temporalmente ajustados com as definições feitas pelo usuário.

Na Figura 5.7 é possível observar a tela de *Data Management* com a aba da escolha de frequência selecionada. Como a frequência da série temporal utilizada como exemplo é mensal, o seletor não foi modificado. Também é interessante visualizar que o seletor da coluna de dados foi modificado para atender à variável que de fato contém os dados temporais, ali definida como *Value*. Já na Figura 5.8, selecionada a aba da escolha do ano de início da série temporal, foi definido o primeiro ano em que os dados temporais estão existentes no conjunto de dados. É notório que, ao realizar qualquer alteração no ano de início da série temporal, o gráfico da *Viewing Pane* é automaticamente atualizado para o período de escolha, sem a necessidade de quaisquer interação extra do usuário. Por último, na Figura 5.9, é mostrada ativa a aba do período de início da série temporal, que para esse exemplo foi utilizado o padrão devido à série temporal ter início no primeiro mês do ano de 1956.

3.4 *Analysis Tools* (Ferramentas de Análise)

O menu *Analysis Tools* consiste em diversas ferramentas oferecidas para modificação da série temporal original, com o objetivo de melhorar a análise ou simplesmente modificá-la. O primeiro sub-menu, com nome análogo ao do menu, consiste de três campos que oferecem diferentes tipos de transformação dos dados. O primeiro oferece uma transformação funcional nos dados temporais, enquanto o segundo e o terceiro campo oferecem a possibilidade de aplicação de diferenças simples no conjunto de dados. Tanto diferenças simples como sa-

zonais são opções disponíveis. O segundo sub-menu, *Statistical Tests* (Testes Estatísticos), oferece alguns testes estatísticos apropriados para essa etapa de análise da série temporal. São eles: Teste da Raiz Unitária de Phillips-Perron, Teste de Dickey-Fuller Aumentado e Teste de Box-Pierce. O primeiro teste é realizado utilizando a função *PP.test* disponível no pacote *stats*. O segundo teste é feito utilizando a função *adftest* disponível no pacote *tseries*. Já o último teste é realizado com a função *Box.test*, também do pacote *stats*. O Teste de Box-Pierce é oferecido numa formatação um pouco diferente dos demais testes, no sentido que é possível realizá-lo para diferentes *lags* da série temporal. Nesse mesmo painel são exibidas as saídas do teste selecionado. Por último, no *Viewing Pane*, são oferecidas três visualizações gráficas da série temporal. A primeira é uma visualização tradicional do gráfico do caminho amostral da série temporal estudada. A segunda e a terceira, visões das autocorrelações da série temporal. As duas últimas ainda possuem um seletor do número de *lags* que deseja-se visualizar.

Na Figura 5.10 é possível verificar os dois primeiros menus da tela: um relativo a transformação de dados e um relativo à testes estatísticos apropriados para a etapa da modelagem em questão. Embora visualmente a série temporal tenha um aspecto não-estacionário, o resultado pelo Teste de Phillips-Perron apresenta a ausência de raiz unitária na série. Na Figura 5.11 é possível observar o correlograma da série temporal em análise. Observamos um lento decaimento das autocorrelações, o que é um possível indicativo de não-estacionaridade. Já na Figura 5.12 temos o correlograma parcial com alguns picos que podem indicar sazonalidade. Nos correlogramas, é interessante notar que o eixo horizontal do gráfico mostra o domínio da frequência da série temporal em detrimento dos anos. Dessa forma, facilita a visualização de um possível padrão nesses gráficos, no intuito de uma modelagem mais robusta.

Nesse sentido, na Figura 5.13 é mostrado o resultado do teste de Box-Pierce para os 5 primeiros *lags* da série, no intuito de verificar se a série temporal pode ser considerada ruído branco (erro aleatório), o que não acontece. Como a série temporal apresenta um possível aumento de variância ao longo do tempo, na Figura 5.14 é mostrada a tela da transformação de dados com a utilização de uma transformação $f(x) = \log(x)$. O gráfico subsequente da série temporal pode ser visualizado na Figura 5.15. Por último, com o objetivo de retirar uma possível tendência e sazonalidade, foi aplicada uma diferença simples e sazonal de *lag* 12 na série e o resultado é mostrado na Figura 5.16. É possível observar uma diferença significativa da série transformada para a série original.

3.5 *Model Adjustment* (Ajuste do Modelo)

O menu oferece então, a partir dos dados temporais ajustados, diferentes opções de modelagem. Sendo assim, são ofertadas a predição de modelo da família ARIMA utilizando a função *auto.arima* do pacote *forecast*. Também é oferecida uma predição de modelo da família GARCH utilizando uma função desenvolvida pelo autor. Outra implementação é a predição de modelo da família ETS utilizando a função *ets* do pacote *forecast*. O nome do modelo é fornecido em um sub-menu, bem como as estimativas e seus respectivos erros padrões. Caso o usuário não esteja satisfeito com os modelos sugeridos, é possível testar um outro modelo diferente dos preditos, utilizando a seleção *User defined* (Definido pelo usuário) do sub-menu *Model Adjustment* (Ajuste do Modelo). A confirmação do nome do modelo bem como suas respectivas estimativas e erros padrões aparecerão da mesma forma que para os anteriores. Para confirmar o modelo a ser utilizado basta prosseguir à próxima etapa da modelagem que ele gravará automaticamente a opção escolhida.

Dentro desse contexto, para utilizar como comparativo da série original versus série transformada, na Figura 5.18 é possível ver o modelo predito para a família SARIMA. Para a série temporal de estudo, o modelo melhor ajustado pelo algoritmo foi um $\text{SARIMA}(0, 1, 1) \times (0, 1, 2)_{12}$. Na Figura 5.19 tem-se o melhor modelo ajustado para a família GARCH. Neste, o melhor modelo encontrado foi um $\text{GARCH}(1, 0)$. Já na Figura 5.20 temos o melhor ajuste para modelo ETS, que no caso foi um $\text{ETS}(M, A, M)$, ou seja, possui erros multiplicativos, tendência aditiva e sazonalidade multiplicativa. Por último, ainda para a série original é mostrado na Figura 5.21 um exemplo modelo definido pelo usuário, através de um sistema de identificação automático que analisa o que foi digitado e modela utilizando função apropriada.

Em contrapartida, para a série transformada, como podemos verificar na Figura 5.22, o modelo selecionado agora foi um $\text{SARIMA}(0, 0, 1) \times (2, 0, 2)_{12}$. Para a família GARCH, como pode ser visualizado na Figura 5.23 não houve mudanças estruturais no modelo encontrando, mantendo-se um $\text{GARCH}(1, 0)$, obviamente com parâmetros diferentes por causa das transformações realizadas nos dados. Na Figura 5.24, o resultado para o modelo ETS foi bastante diferente para a série transformada em comparação com a original, ofertando um modelo $\text{ETS}(A, N, N)$ estatisticamente pobre em relação ao selecionado para a série original. Isso ocorre pelo fato da modelagem exponencial não ter bom desempenho quando os dados são estacionários.

3.6 *Residual Analysis* (Análise de Resíduos)

Após a definição do modelo a ser utilizado, uma análise dos resíduos do modelo é oferecida no objetivo de verificar alguns pressupostos existentes. É oferecido também outros testes estatísticos apropriados para verificação dos resíduos como: Teste de Shapiro-Wilk, Teste de Breusch-Pagan e Teste de Box-Pierce. O primeiro teste é realizado utilizando a função *shapiro.test* do pacote *stats*. O segundo teste é feito utilizando a função *bptest* do pacote *lmtest* e o último já foi anteriormente definido na seção 3.5. O último sub-menu oferece a possibilidade de visualização da série temporal dos resíduos, bem como o correlograma e o correlograma parcial dos resíduos no objetivo de verificar se o modelo foi adequadamente ajustado.

Com o objetivo de uma visualização abrangente dos resíduos, pode-se verificar para a figura 5.25 o gráfico Q-Q *Plot*². Na Figura 5.26 temos um histograma dos dados aderido à uma distribuição normal, com o objetivo de se comparar visualmente se o ajuste à normalidade ocorre no caso. Na Figura 5.28 temos a série dos resíduos com alguns *p*-valores obtidos através do teste de Box-Pierce. Note que todos foram superiores a 0.05, ou seja, é possível inferir que a 95% de confiança os resíduos não estão correlacionados. Ainda no mesmo âmbito, na Figura 5.27 é possível ver o Diagrama de Cullen e Frey³ para a série dos resíduos. Já nas Figuras 5.29 e 5.30 é possível visualizar os respectivos correlogramas para a série dos resíduos.

3.7 *Model Forecast* (Previsão do Modelo)

O menu oferece a possibilidade do usuário então verificar a qualidade de seu ajuste. O primeiro sub-menu, *Forecasting Options* (Opções de Previsão), oferece algumas escolhas para definir a previsão, como o número de observações à frente (definido aqui pela frequência em que os dados foram ajustados) e o nível de significância para visualização da taxa de ajustes corretos, como pode ser melhor visualizado na Figura 5.31. No segundo sub-menu, *Forecasting Accuracy* (Acurácia da Previsão) mostra o número de observações ajustadas corretamente

²Q-Q *Plot* (Gráfico Quantil-Quantil) é um gráfico de probabilidade que compara duas distribuições, traçando seus quantis um contra o outro. (GNANADESIKAN, 1968). Neste caso, o objetivo é mostrar a defasagem dos dados residuais em relação à Distribuição Normal.

³O Diagrama de Cullen e Frey é utilizado como uma alternativa possível para determinação da distribuição de probabilidade dos dados, considerando em um dos eixos a curtose e no outro o quadrado da assimetria. (CULLEN e FREY, 1999).

ao nível de significância definido no sub-menu anterior. Em *Forecast Estimates* (Estimativas da Previsão) é possível visualizar os valores futuros preditos, bem como seus intervalos de confiança. Por último, é possível visualizar a série temporal junto com os intervalos de confiança da previsão para o futuro, definido no número de passos escolhidos no primeiro sub-menu. Na Figura 5.32 temos as estimativas pontuais e a 95% de confiança para 10 passos à frente, conforme previamente selecionado. Já na Figura 5.33 temos o gráfico da série temporal conjuntamente com os dados da previsão (em azul) e suas bandas de confiança (em cinza). Em termos de comparação, podemos verificar na Figura 5.34 a mesma ideia, porém com previsão de 100 passos à frente. Note que o intervalo de confiança não é estático e vai aumentando sua amplitude conforme a distância da observação a ser predita. Na Figura 5.35 há uma recapitulação do menu *Model Adjustment*, onde definiu-se a “reciclagem” de 10 observações com o objetivo de utilizar para cálculo da acurácia do modelo, conforme pode ser visto na Figura 5.36, onde foi encontrada uma taxa de 80% de acertos para as últimas 10 observações da série, a nível de 95% de confiança.

3.8 *Plot Options* (Opções Gráficas)

O menu tem como objetivo dar o usuário uma aparência mais pessoal a seus gráficos, possibilitando tanto a alteração do título do gráfico quanto da cor através de uma paleta de cores hexadecimal. Nas Figuras 5.37, 5.38 e 5.39 é possível visualizar, respectivamente, as opções de customização disponíveis, a aparência da paleta de cores focada na cor violeta e o gráfico da série temporal com a nova cor escolhida.

3.9 *Generate Report* (Geração de Relatório)

O menu oferece a possibilidade do usuário realizar o *download* de um relatório com os principais gráficos, tabelas e saídas pertinentes à análise realizada na interface, conforme pode ser visualizado na Figura 5.40.

Capítulo 4

Considerações Finais

Este trabalho, que teve como objetivo o desenvolvimento de uma ferramenta que possa ter utilidade global, tem sua distribuição online conforme mostrado no Capítulo 3 e, futuramente, haverá realização de melhorias para que a interface gráfica cada vez esteja num melhor nível, podendo haver alterações em relação ao *layout* original.

Além disso, é notável a percepção do valor útil do conceito análise de séries temporais em praticamente qualquer área do conhecimento, o que instigou também o desenvolvimento do projeto.

E também, com o crescente desenvolvimento da linguagem de programação R, a expectativa é que os horizontes se expandam na área de interfaces gráficas.

Capítulo 5

Anexos

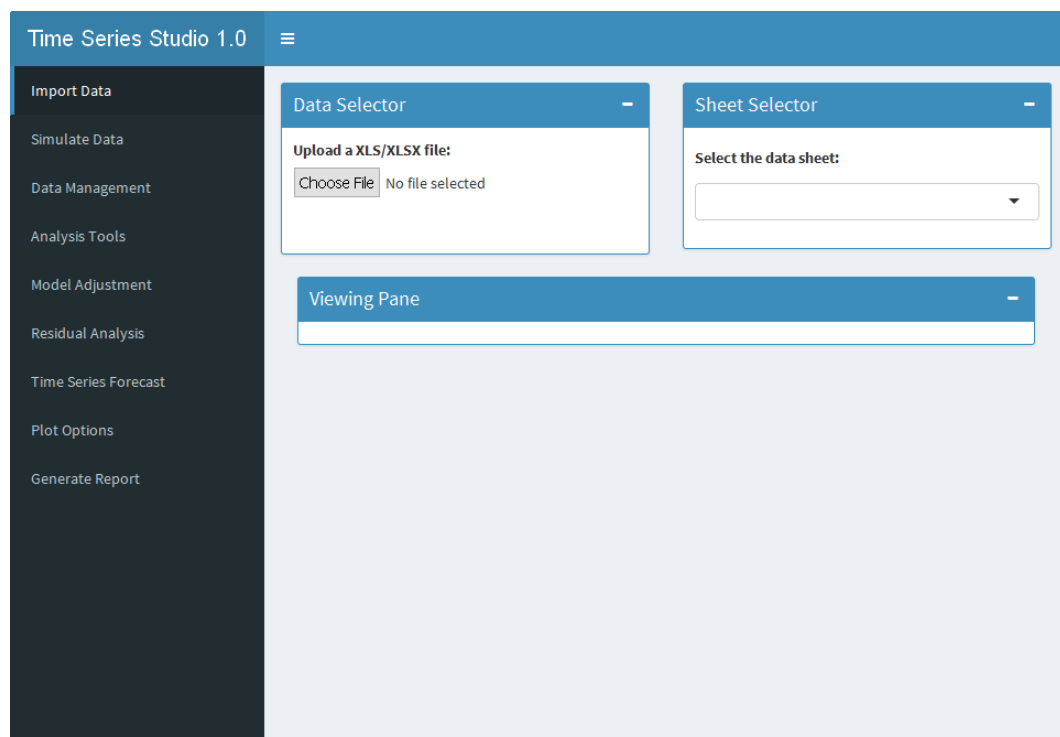


Figura 5.1: Tela inicial do programa

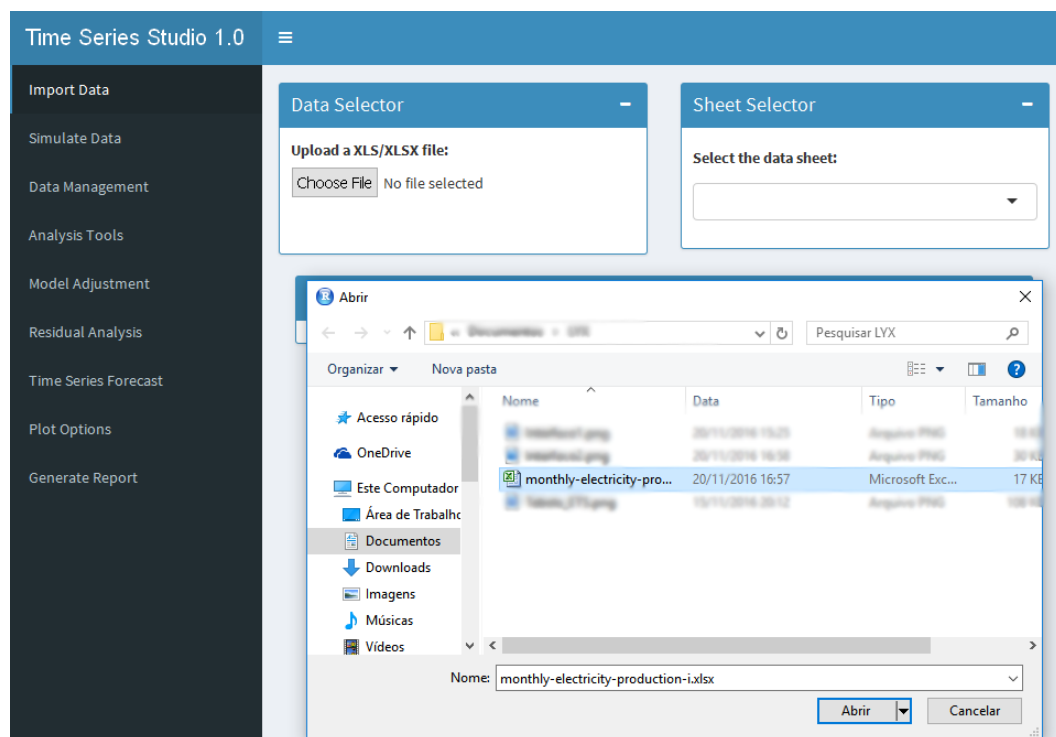


Figura 5.2: Tela após o botão do *Data Selector* ser ativado

Time Series Studio 1.0

Import Data

Simulate Data

Data Management

Analysis Tools

Model Adjustment

Residual Analysis

Time Series Forecast

Plot Options

Generate Report

Data Selector

Upload a XLS/XLSX file:

Choose File ... tricity-production-1.xlsx

Upload complete

Sheet Selector

Select the data sheet:

Monthly electricity production

Viewing Pane

Show 10 entries

Search:

	Date	Value
1	1956-01-01	1254
2	1956-02-01	1290
3	1956-03-01	1379
4	1956-04-01	1346
5	1956-05-01	1535
6	1956-06-01	1555
7	1956-07-01	1655

Showing 1 to 10 of 476 entries

Previous 1 2 3 4 5 ... 48 Next

Figura 5.3: Tela após a importação dos dados

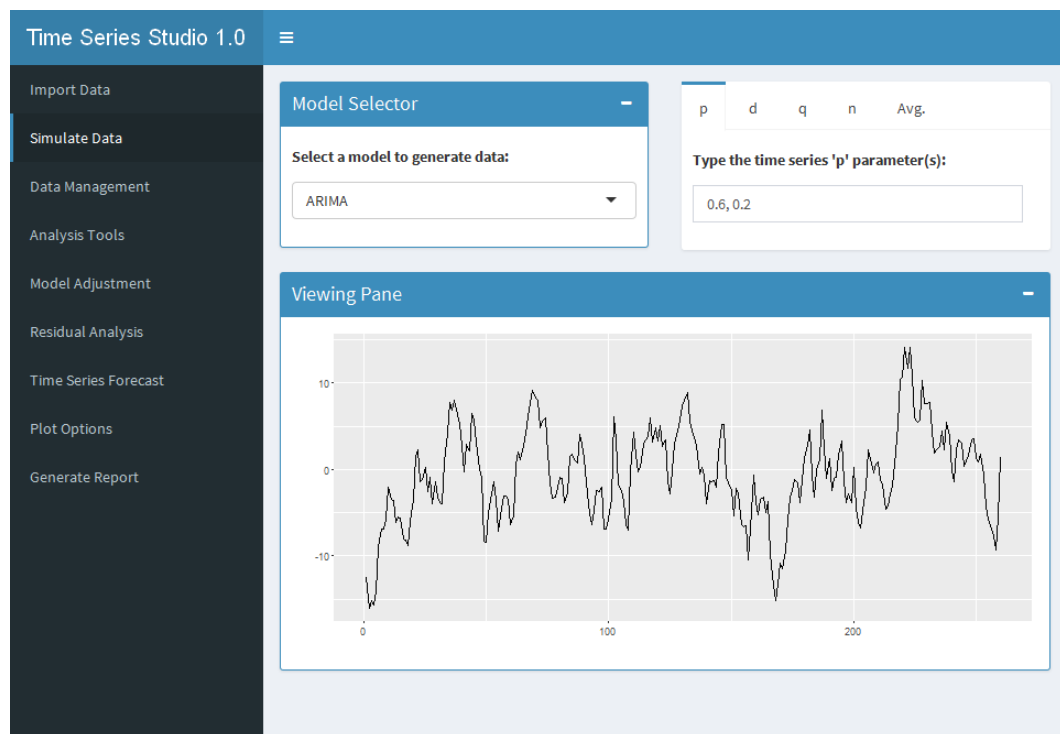


Figura 5.4: Tela de *Simulate Data* com a geração de série através de Modelo ARIMA

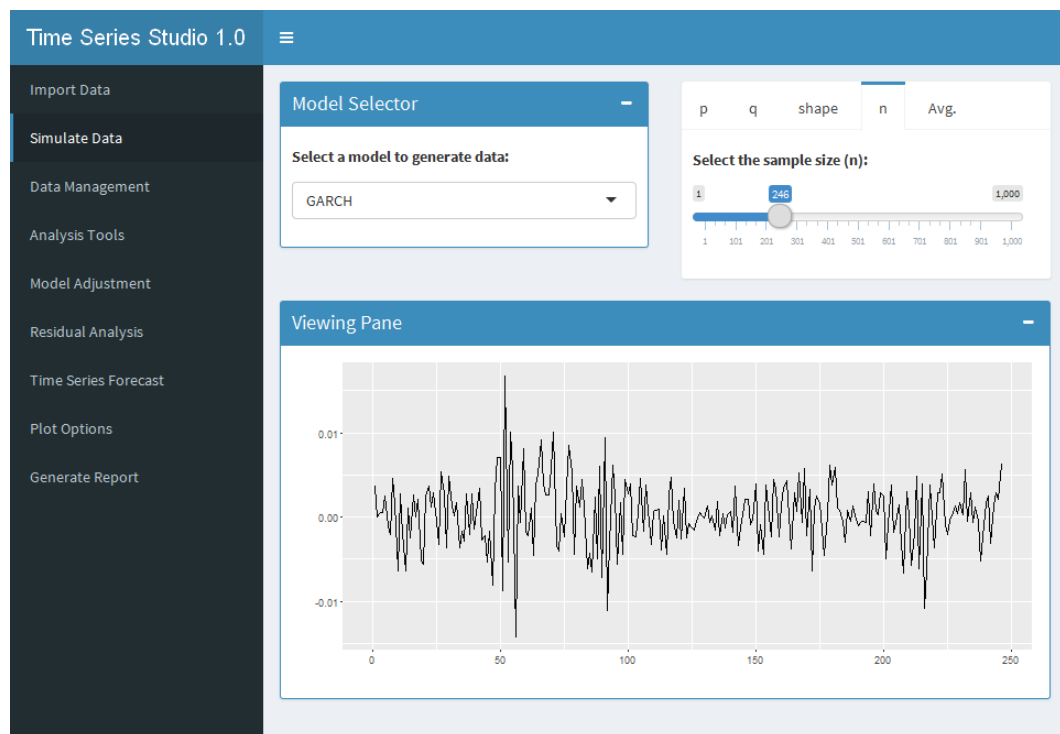


Figura 5.5: Tela de *Simulate Data* com a geração de série através de Modelo GARCH

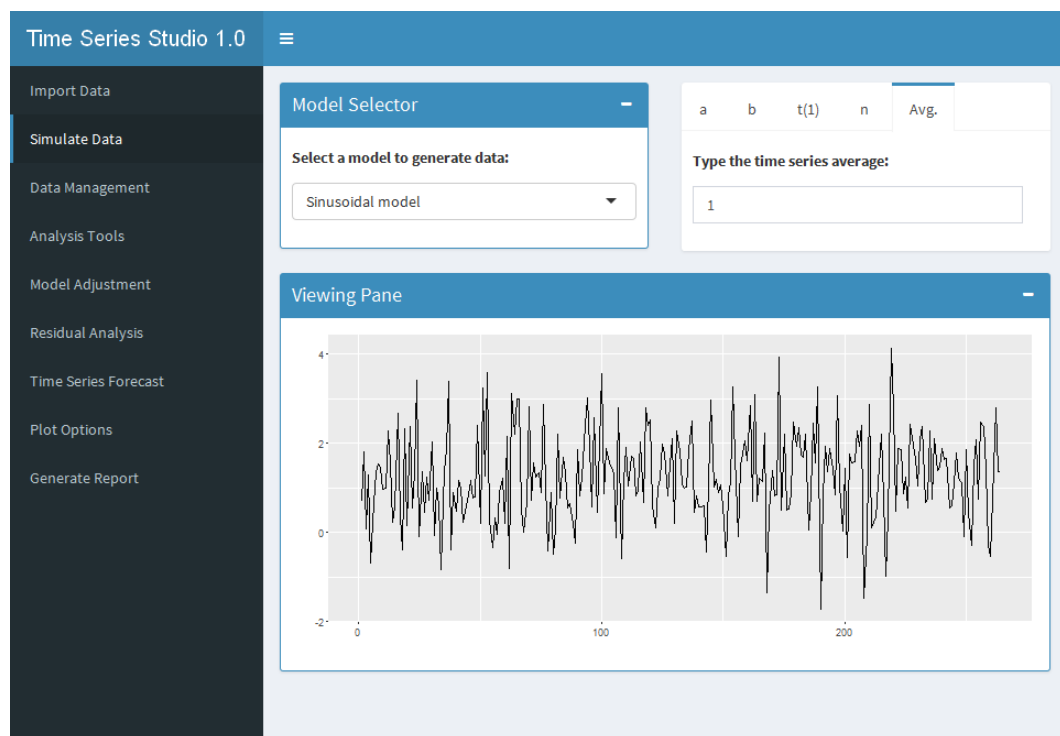


Figura 5.6: Tela de *Simulate Data* com a geração de série através de Modelo Senoidal

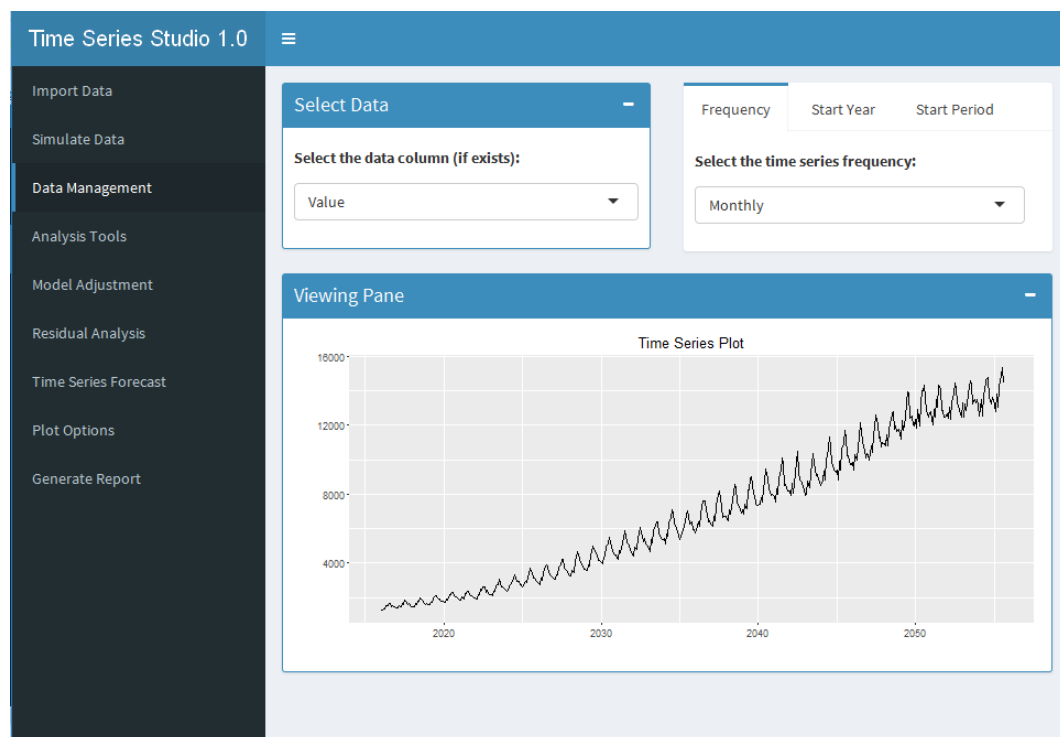


Figura 5.7: Tela de *Data Management* com a aba *Frequency* selecionada

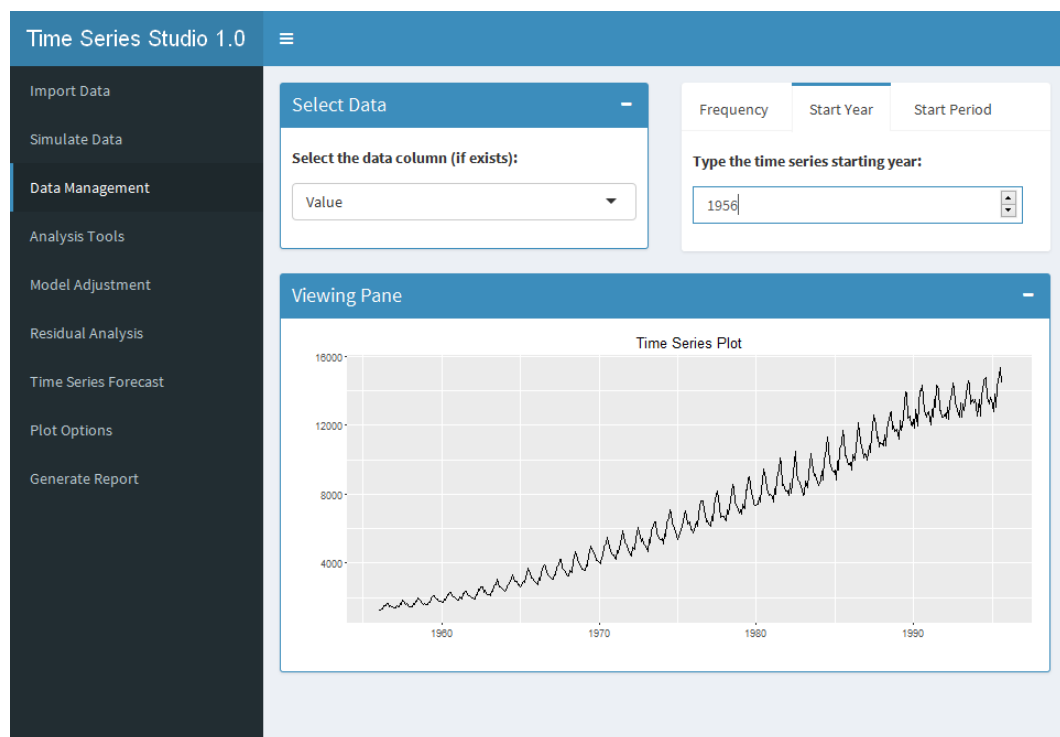


Figura 5.8: Tela de *Data Management* com a aba *Start Year* selecionada

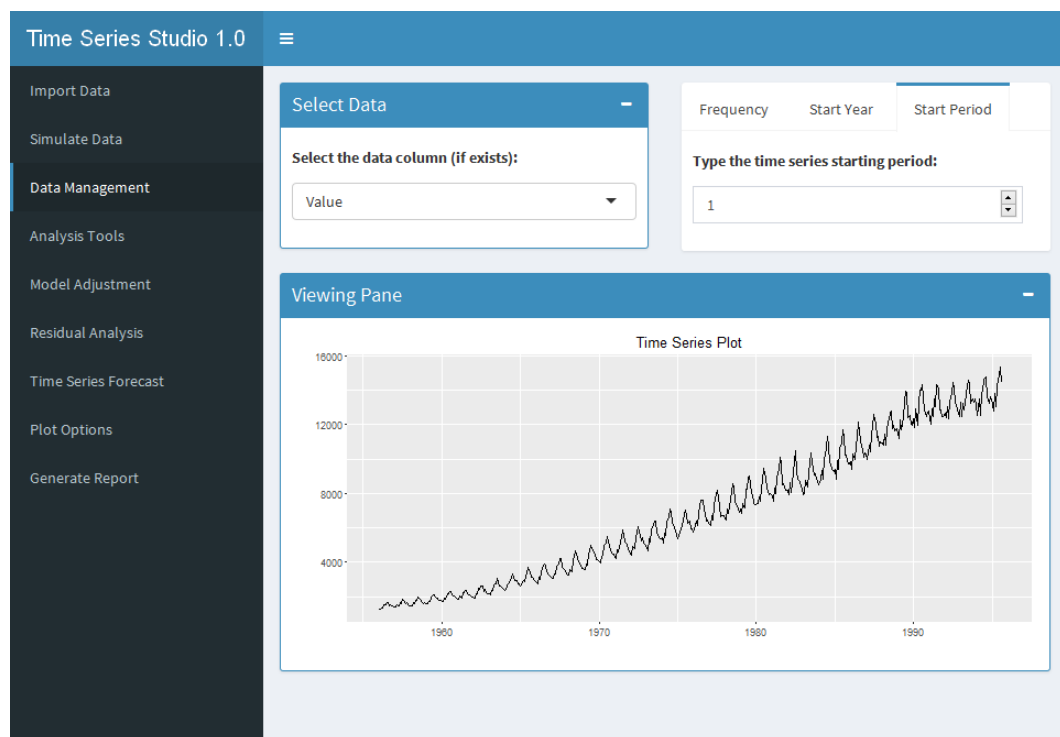


Figura 5.9: Tela de *Data Management* com a aba *Start Period* selecionada

Import Data

Simulate Data

Data Management

Analysis Tools

Model Adjustment

Residual Analysis

Time Series Forecast

Plot Options

Generate Report

Analysis Tools

Apply a $f(x)$ transformation:

Apply simple differences:

0

Apply seasonal differences based on frequency:

0

Statistical Tests

Select a statistical test:

Phillips-Perron Unit Root Test

Show 10 entries

Search:

	Statistic	Value
1	Dickey-Fuller Statistic	-7.5108
2	p-value	0.0100

Showing 1 to 2 of 2 entries

Previous1Next

Figura 5.10: Tela de *Analysis Tools* mostrando a tela de transformação de dados e testes estatísticos, este selecionado no Teste de Phillips-Perron

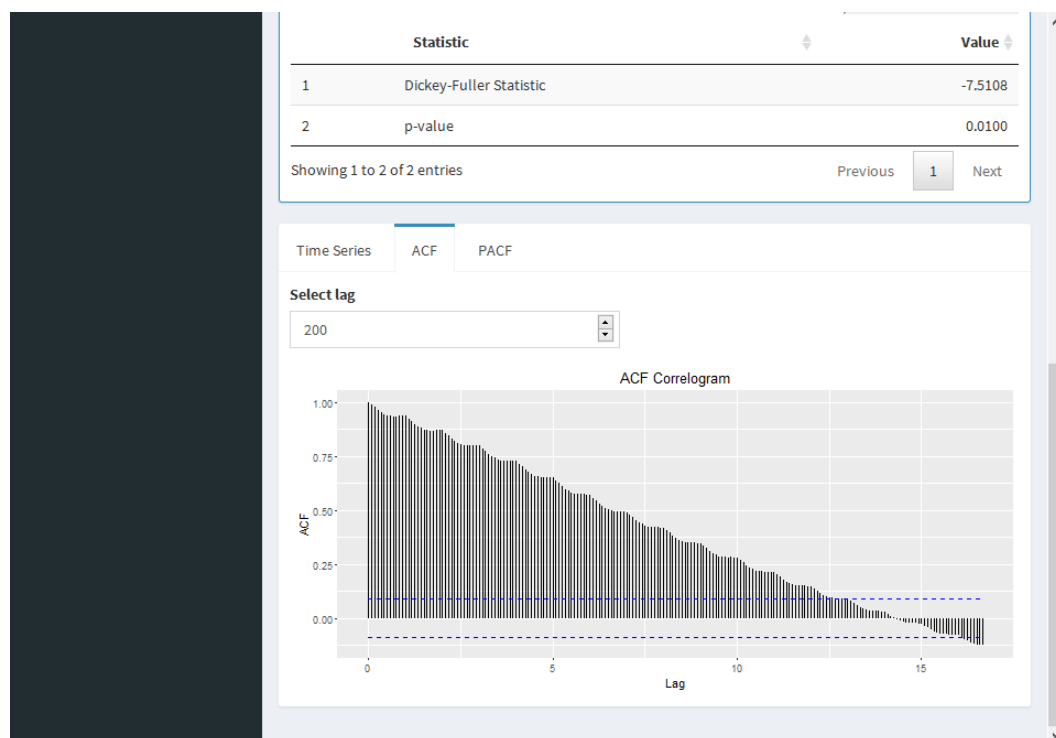


Figura 5.11: Tela de *Analysis Tools* mostrando o correlograma numa amplitude de 200 lags

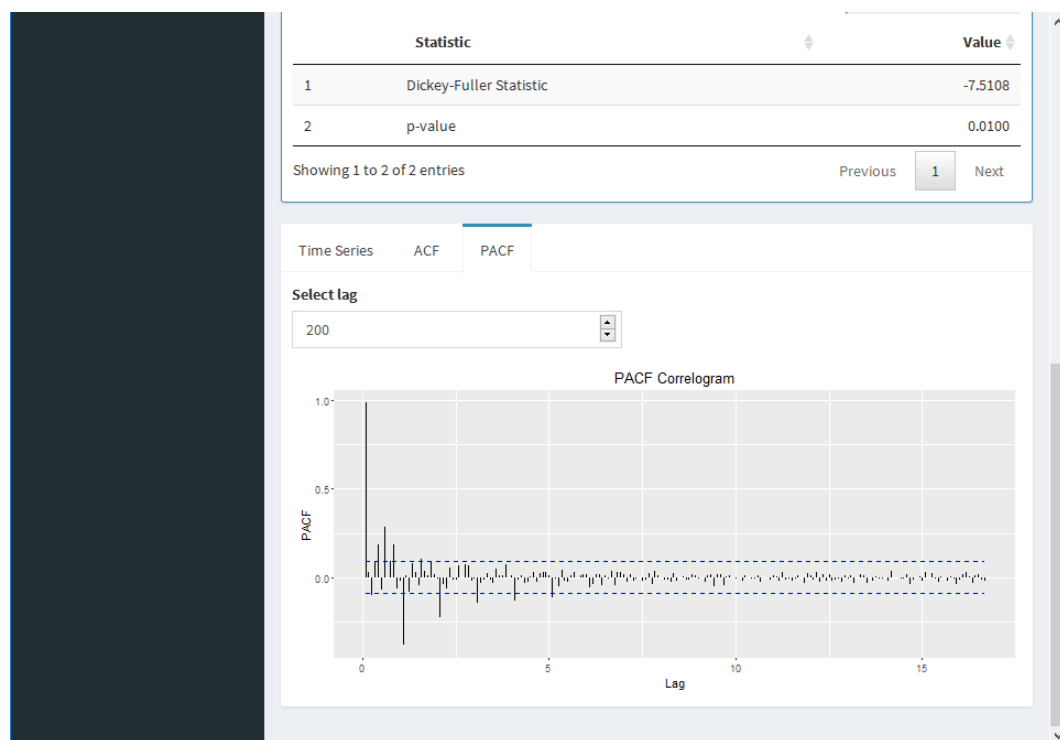


Figura 5.12: Tela de *Analysis Tools* mostrando o correlograma parcial numa amplitude de 200 lags

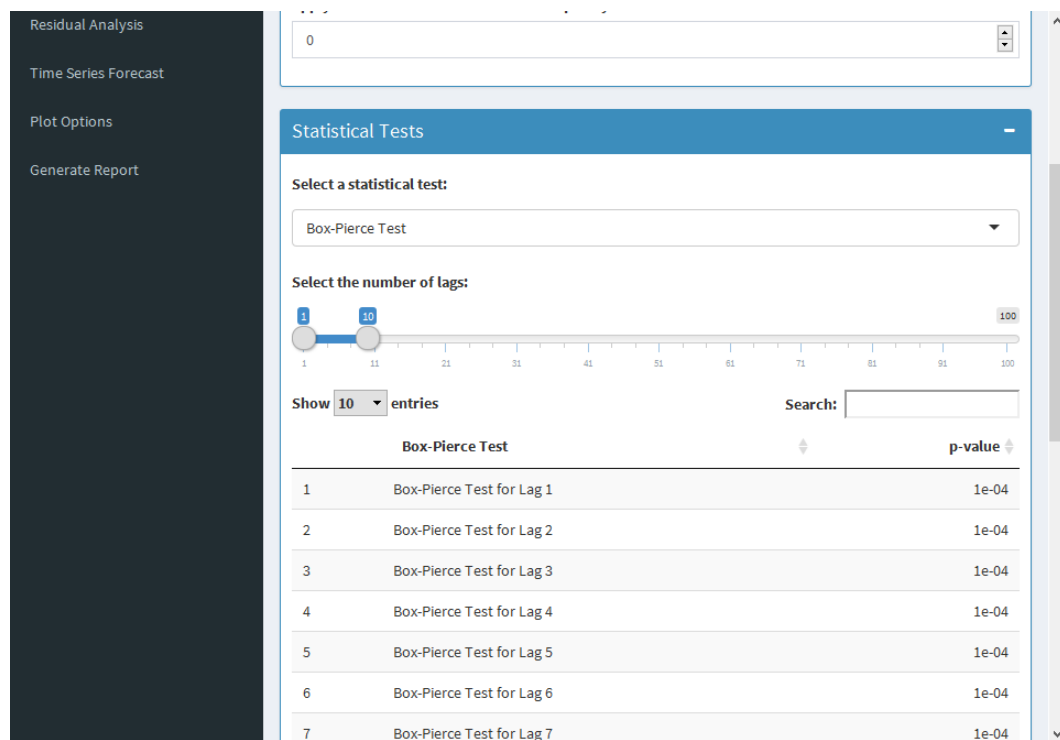


Figura 5.13: Tela de *Analysis Tools* mostrando o resultado do teste de Box-Pierce para os 5 primeiros *lags*

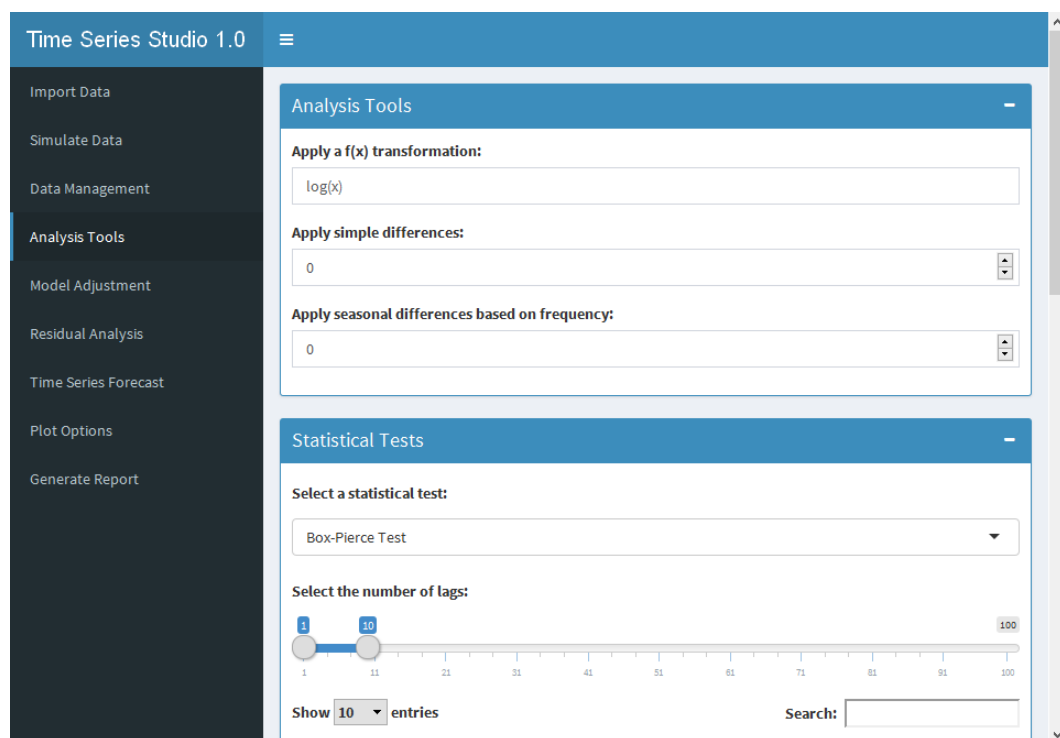


Figura 5.14: Tela de *Analysis Tools* mostrando a aplicação de uma transformação $f(x) = \log(x)$

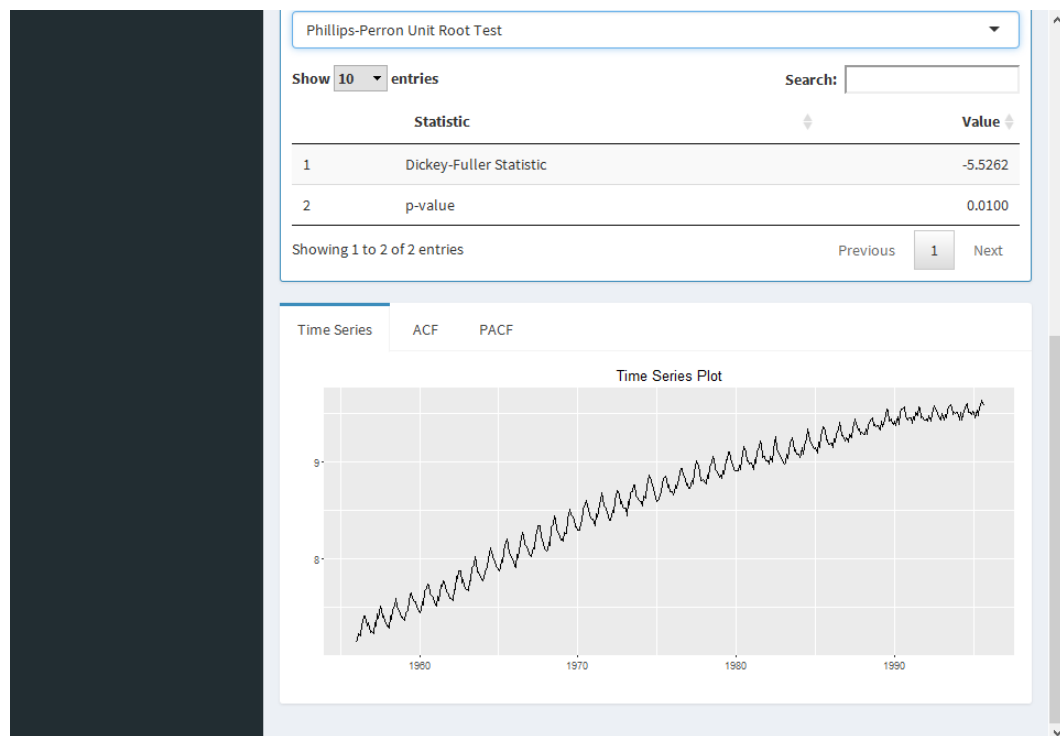


Figura 5.15: Tela de *Analysis Tools* mostrando a nova série temporal transformada

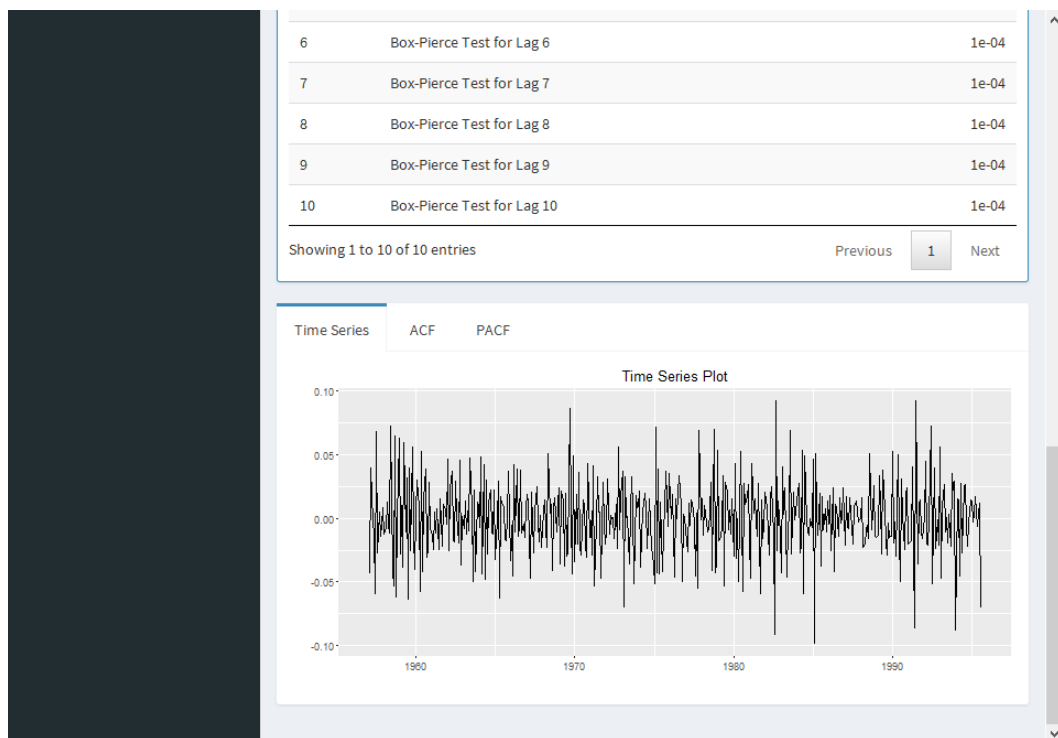


Figura 5.16: Tela de *Analysis Tools* mostrando a série temporal transformada após uma diferença simples e uma sazonal

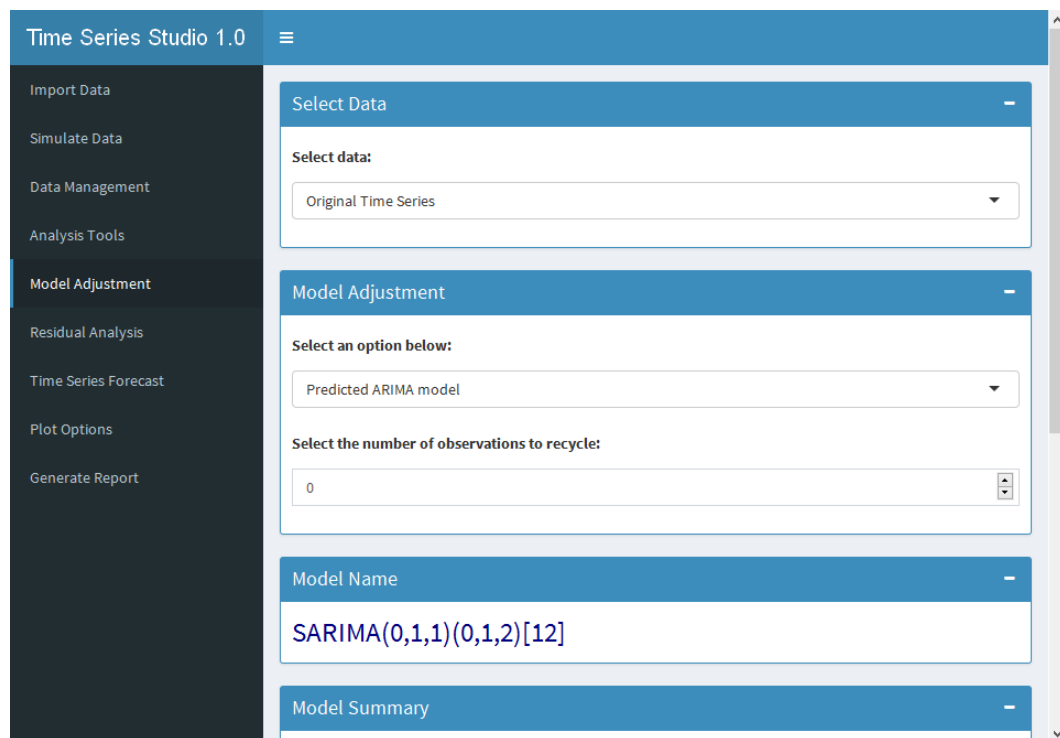


Figura 5.17: Tela de *Model Adjustment* com os dados temporais originais selecionados e o modelo predito da família SARIMA

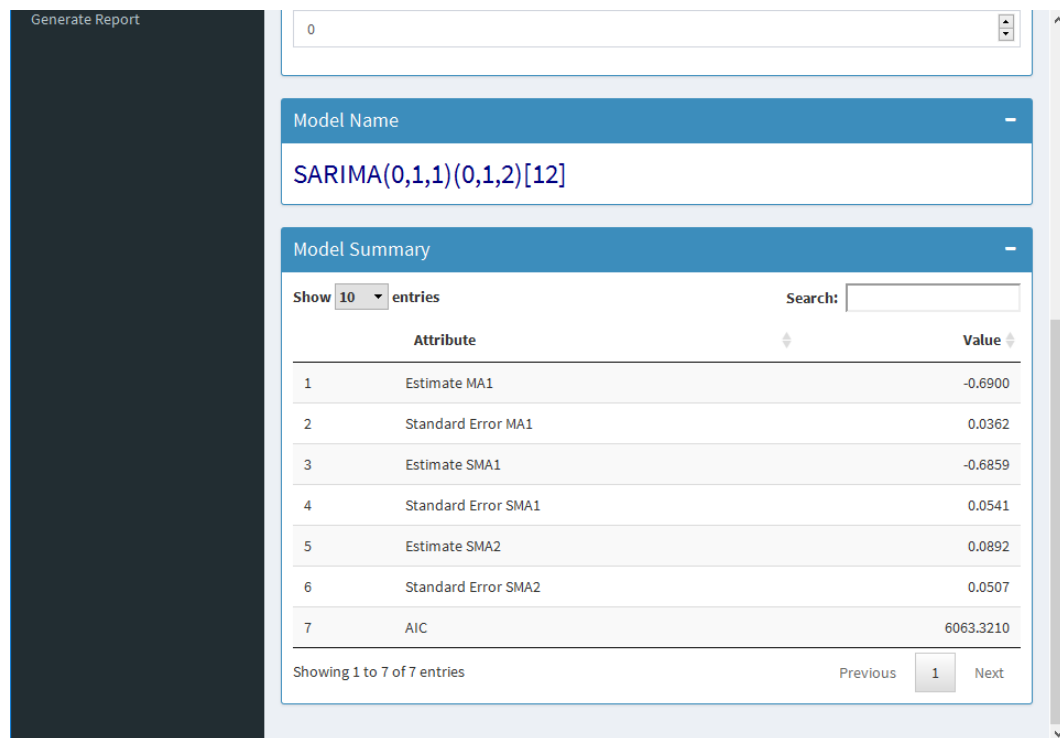


Figura 5.18: Tela de *Model Adjustment* com o modelo predito da família SARIMA para a série original, suas estimativas, erros padrões e AIC

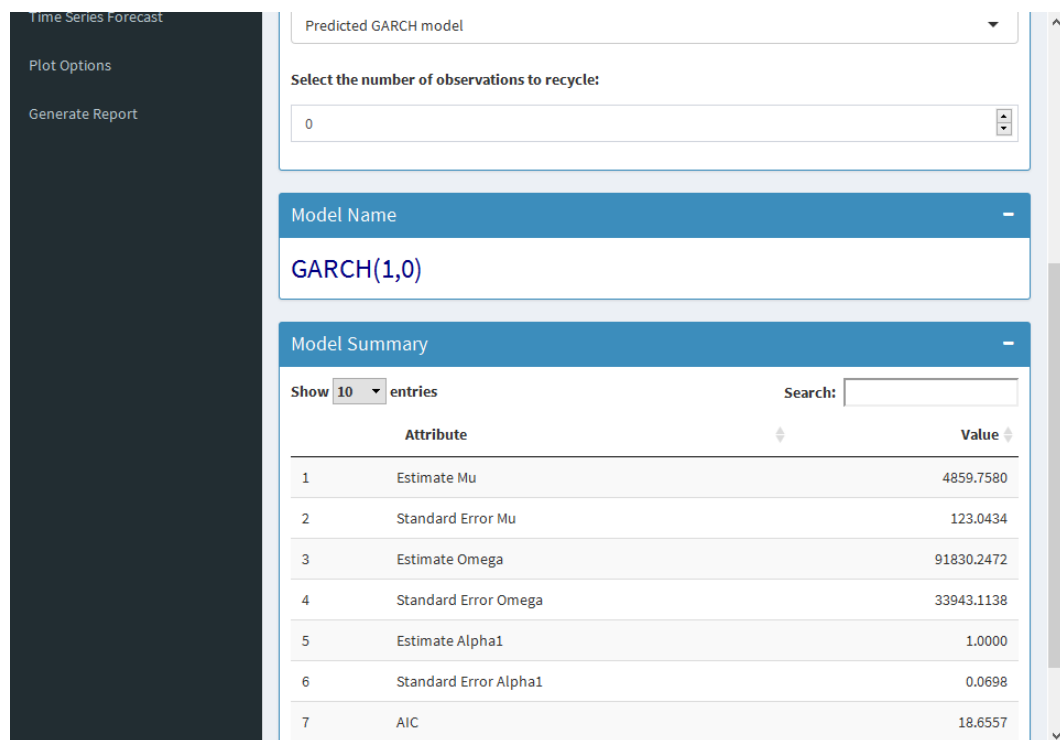


Figura 5.19: Tela de *Model Adjustment* com o modelo predito da família GARCH para série original, suas estimativas, erros padrões e AIC

ETS(M,A,M)

Model Summary

Show 10 entries

Search:

	Attribute	Value
1	Estimate alpha	0.2285
2	Estimate beta	0.0086
3	Estimate gamma	0.3504
4	Estimate l	1403.5252
5	Estimate b	11.9839
6	Estimate s0	0.9146
7	Estimate s1	0.9642
8	Estimate s2	1.0051
9	Estimate s3	1.0039
10	Estimate s4	1.1013

Showing 1 to 10 of 17 entries

Previous

1

2

Next

Figura 5.20: Tela de *Model Adjustment* com o modelo predito da família ETS para a série original, suas estimativas, erros padrões e AIC

Model Adjustment
Residual Analysis
Time Series Forecast
Plot Options
Generate Report

Model Adjustment

Select an option below:

User defined

Define a model (eg.: ARIMA(1,0,1))

SARIMA(1,0,0)(1,0,0)

Select the number of observations to recycle:

0

Model Name

SARIMA(1,0,0)(1,0,0)[12]

Model Summary

Show 10 entries
Search:

	Attribute	Value
1	Estimate AR1	0.9333
2	Standard Error AR1	0.0247

Figura 5.21: Tela de *Model Adjustment* com um modelo SARIMA definido pelo usuário para a série original, suas estimativas, erros padrões e AIC

Simulate Data

Data Management

Analysis Tools

Model Adjustment

Residual Analysis

Time Series Forecast

Plot Options

Generate Report

Select Data

Select data:

Transformed Time Series

Model Adjustment

Select an option below:

Predicted ARIMA model

Select the number of observations to recycle:

0

Model Name

SARIMA(0,0,1)(2,0,2)[12]

Model Summary

Show 10 entries

Search:

Attribute	Value
-----------	-------

Figura 5.22: Tela de *Model Adjustment* com o modelo predito da família SARIMA para a série transformada, suas estimativas, erros padrões e AIC

Simulate Data
Data Management
Analysis Tools
Model Adjustment
Residual Analysis
Time Series Forecast
Plot Options
Generate Report

Select data:
Transformed Time Series

Model Adjustment

Select an option below:
Predicted GARCH model

Select the number of observations to recycle:
0

Model Name
GARCH(1,0)

Model Summary

Show 10 entries
Search:

	Attribute	Value
1	Estimate Mu	-0.0001

Figura 5.23: Tela de *Model Adjustment* com o modelo predito da família GARCH para a série transformada, suas estimativas, erros padrões e AIC

Residual Analysis

Time Series Forecast

Plot Options

Generate Report

Select an option below:

Predicted ETS model

Select the number of observations to recycle:

0

Model Name

ETS(A,N,N)

Model Summary

Show 10 entries

Search:

	Attribute	Value
1	Estimate alpha	0.0001
2	Estimate l	-0.0003
3	AIC	-397.0474

Showing 1 to 3 of 3 entries

Previous1Next

Figura 5.24: Tela de *Model Adjustment* com o modelo predito da família ETS para a série transformada, suas estimativas, erros padrões e AIC

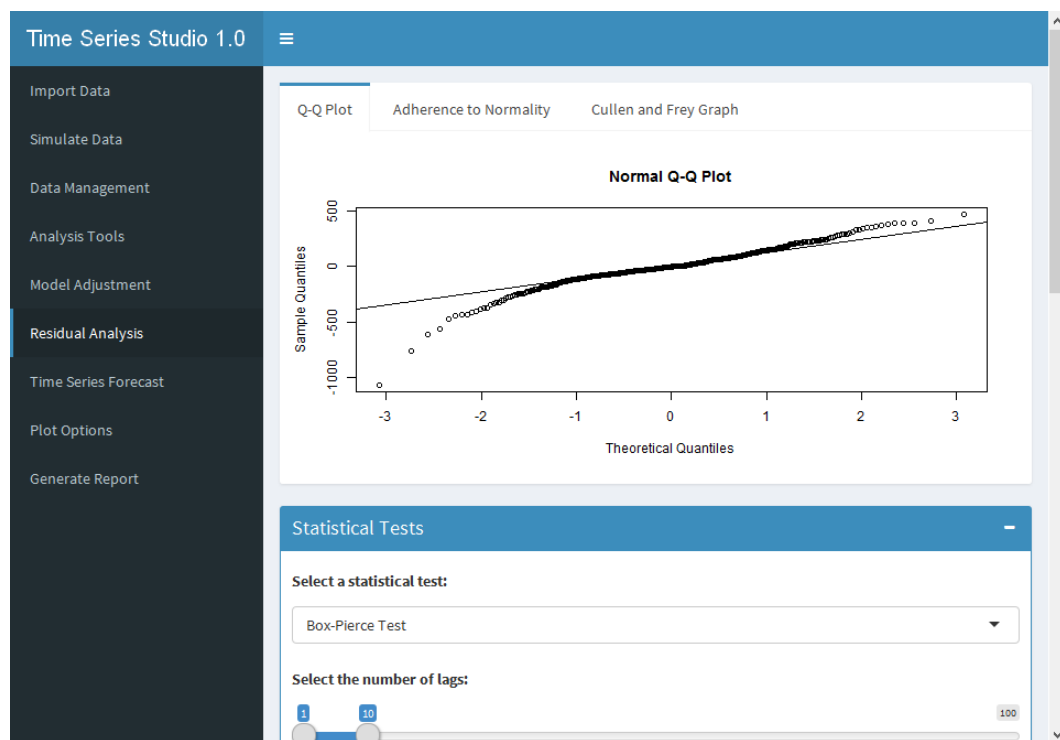


Figura 5.25: Tela de Residual Analysis com o Q-Q Plot

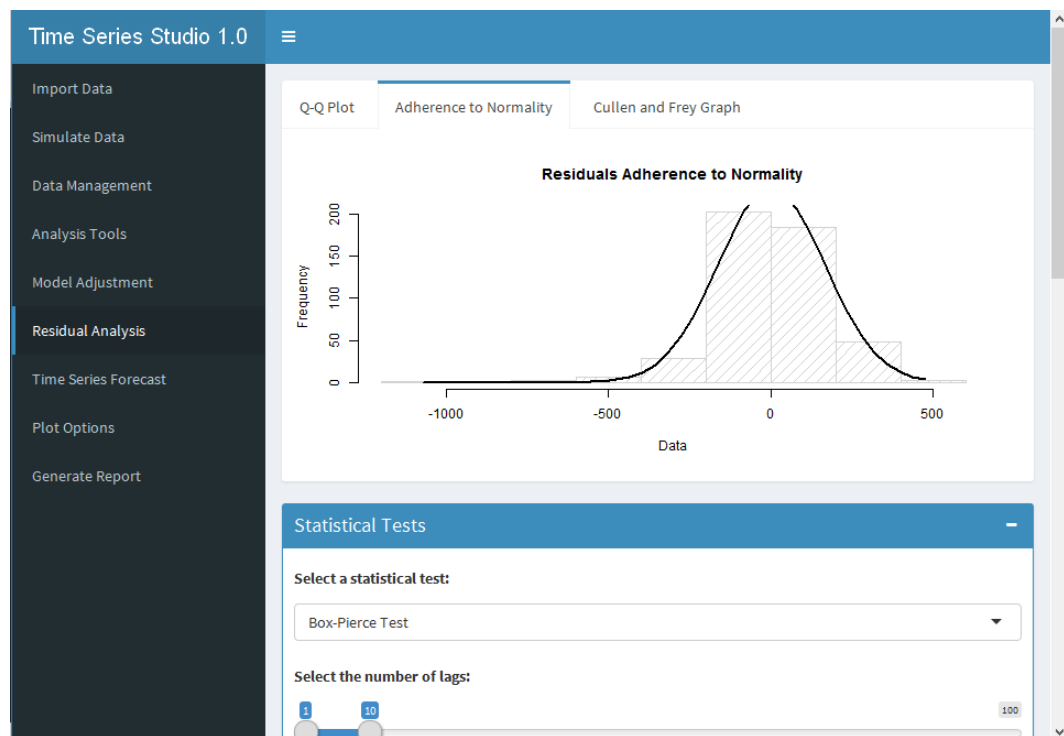


Figura 5.26: Tela de *Model Adjustment* com o histograma da aderência dos resíduos à distribuição normal

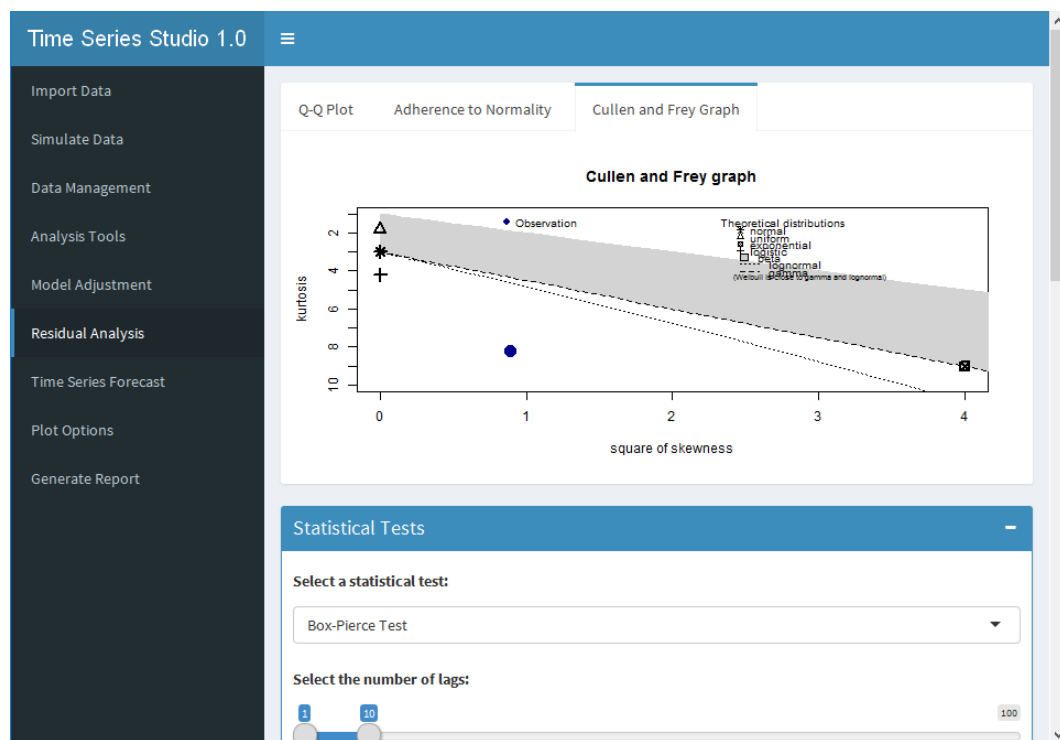


Figura 5.27: Tela de *Residual Analysis* com o Diagrama de Cullen e Frey

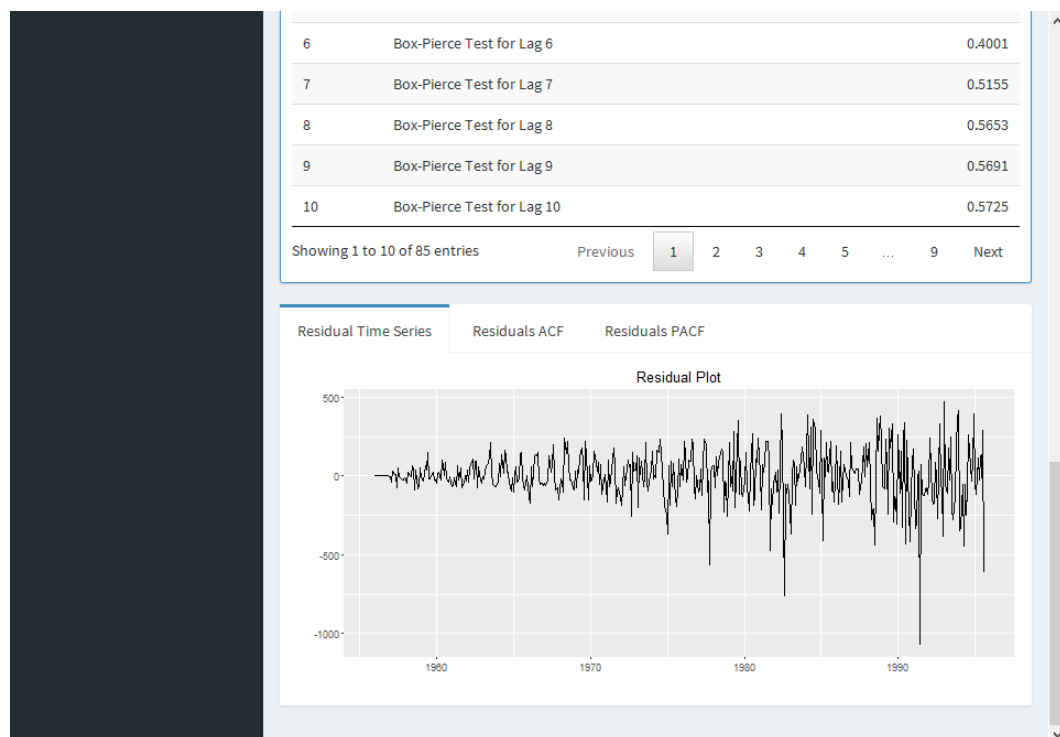


Figura 5.28: Tela de *Residual Analysis* com a série dos resíduos

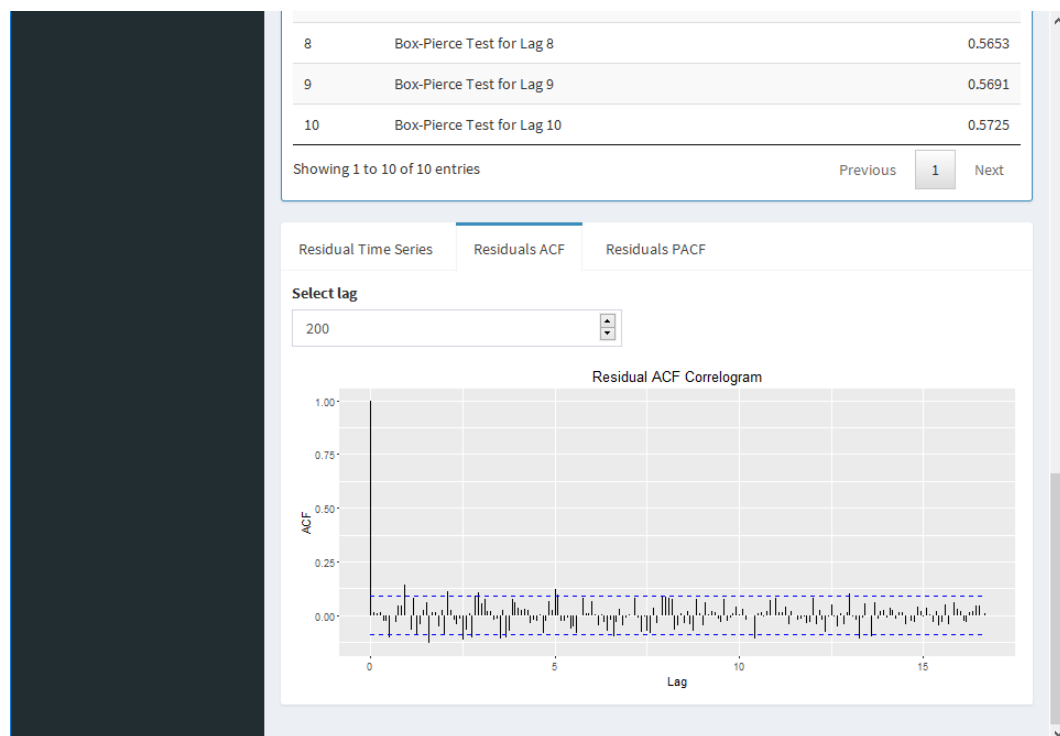


Figura 5.29: Tela de *Residual Analysis* com o correlograma da série dos resíduos com uma amplitude de 200 lags

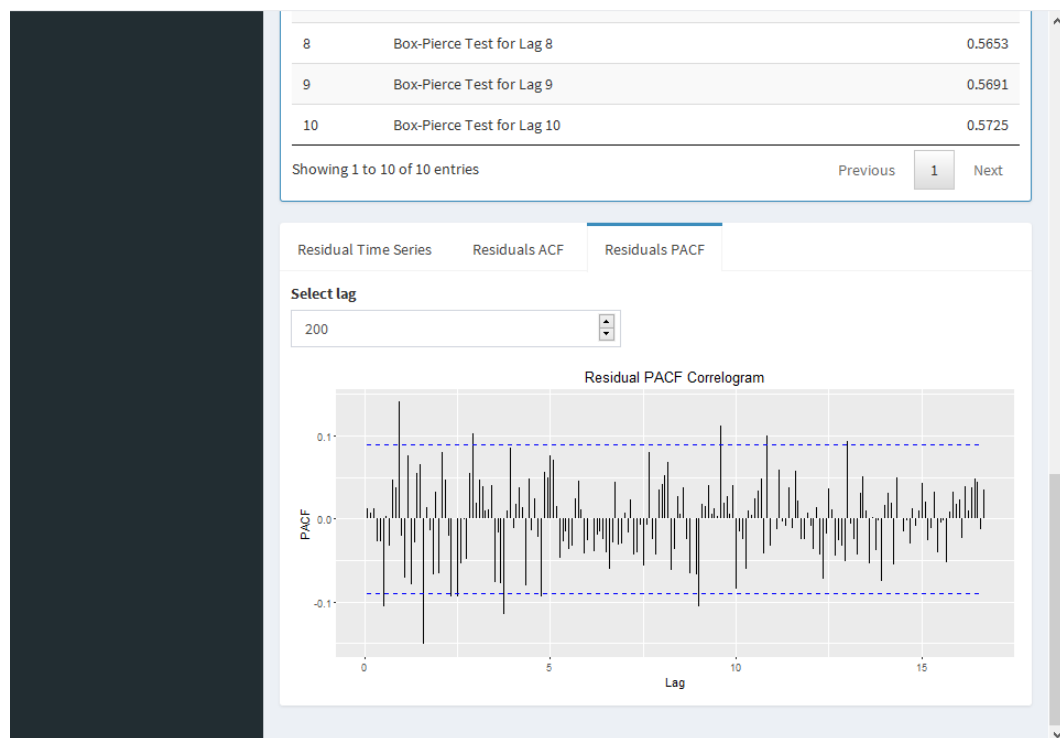


Figura 5.30: Tela de *Residual Analysis* com o correlograma parcial da série dos resíduos com uma amplitude de 200 lags

Time Series Studio 1.0

Import Data

Simulate Data

Data Management

Analysis Tools

Model Adjustment

Residual Analysis

Time Series Forecast

Plot Options

Generate Report

Forecasting Options

Select the steps forward to forecasting:

10

Type the significance level (%):

5

Forecasting Accuracy

Forecasting Estimates

Show 10 entries

Search:

	Point Forecast	Lo 95	Hi 95
Sep 1995	13758.87	13432.19	14085.56
Oct 1995	13753.85	13411.83	14095.88
Nov 1995	13479.82	13123.11	13836.53

Figura 5.31: Tela de *Time Series Forecast* com as opções iniciais de número de passos para previsão e nível de significância

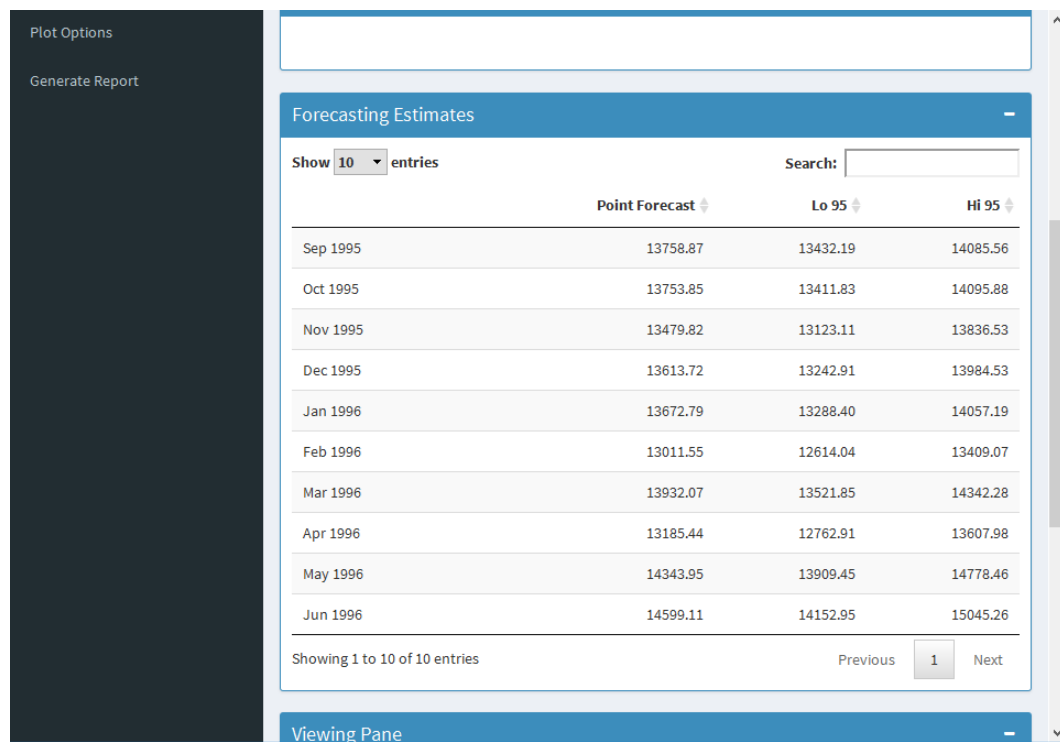


Figura 5.32: Tela de *Time Series Forecast* com as 10 primeiras estimativas pontuais e intervalar a 95% de confiança

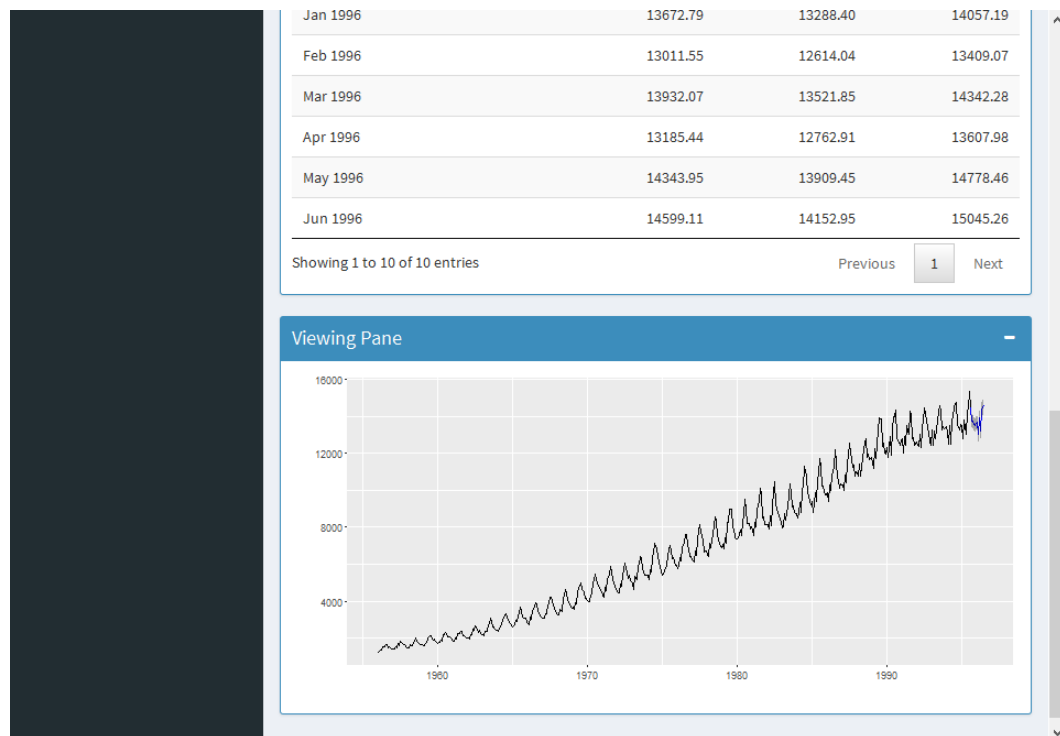


Figura 5.33: Tela de *Time Series Forecast* com o gráfico da série original adicionado da previsão 10 passos à frente junto aos seus respectivos intervalos de confiança a 95%

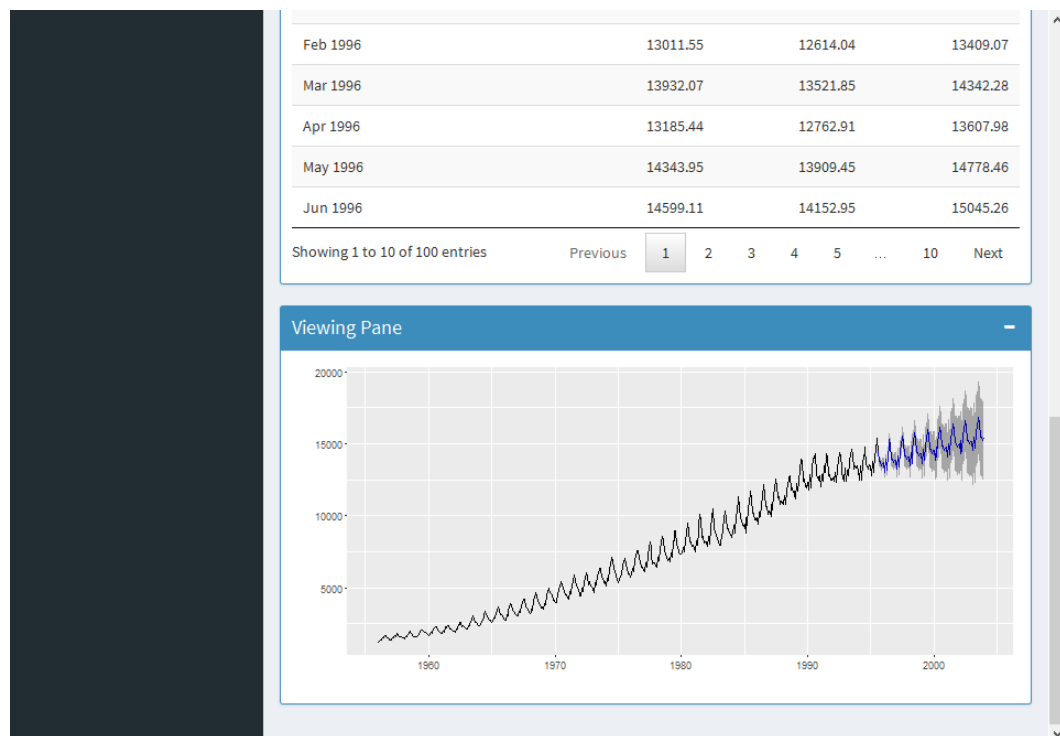


Figura 5.34: Tela de *Time Series Forecast* com o gráfico da série original adicionado da previsão 100 passos à frente junto aos seus respectivos intervalos de confiança a 95%

Import Data

Simulate Data

Data Management

Analysis Tools

Model Adjustment

Residual Analysis

Time Series Forecast

Plot Options

Generate Report

Select Data

Select data:

Original Time Series

Model Adjustment

Select an option below:

Predicted ARIMA model

Select the number of observations to recycle:

10

Model Name

SARIMA(0,1,1)(0,1,2)[12]

Model Summary

Show 10 entries

Search:

Attribute

Value

Figura 5.35: Tela de *Model Adjustment* “reciclando” as 10 últimas observações da série com o objetivo de testar a previsão do modelo

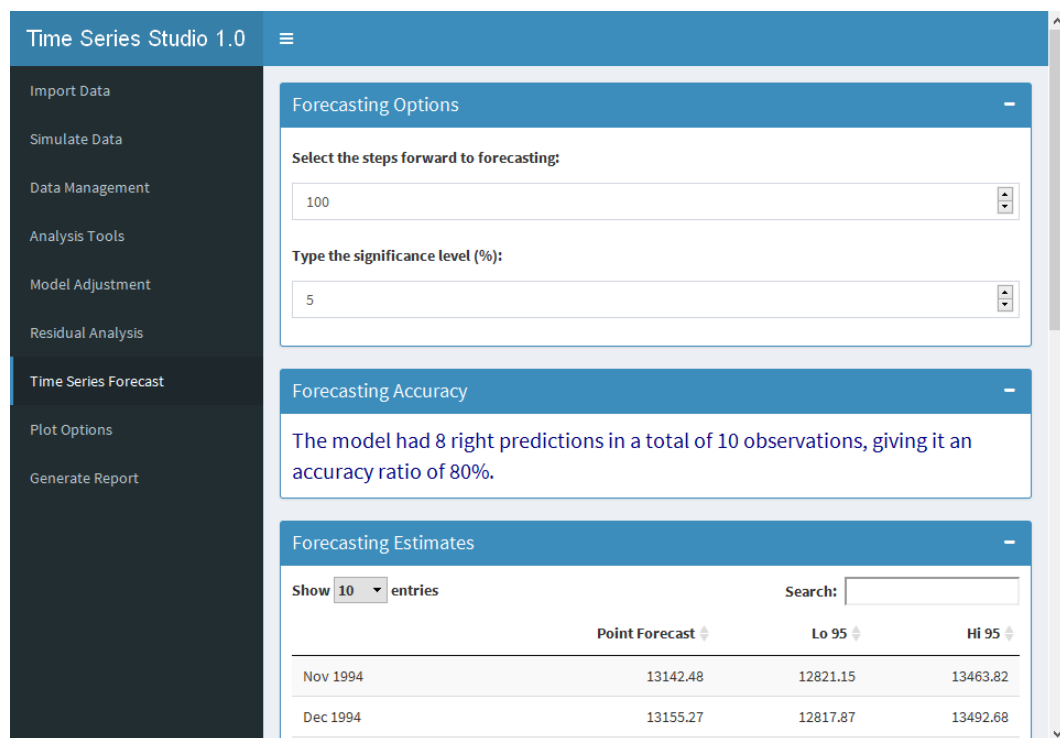


Figura 5.36: Tela de *Time Series Forecast* apresentando a acurácia da previsão baseado nos dados “reciclados”

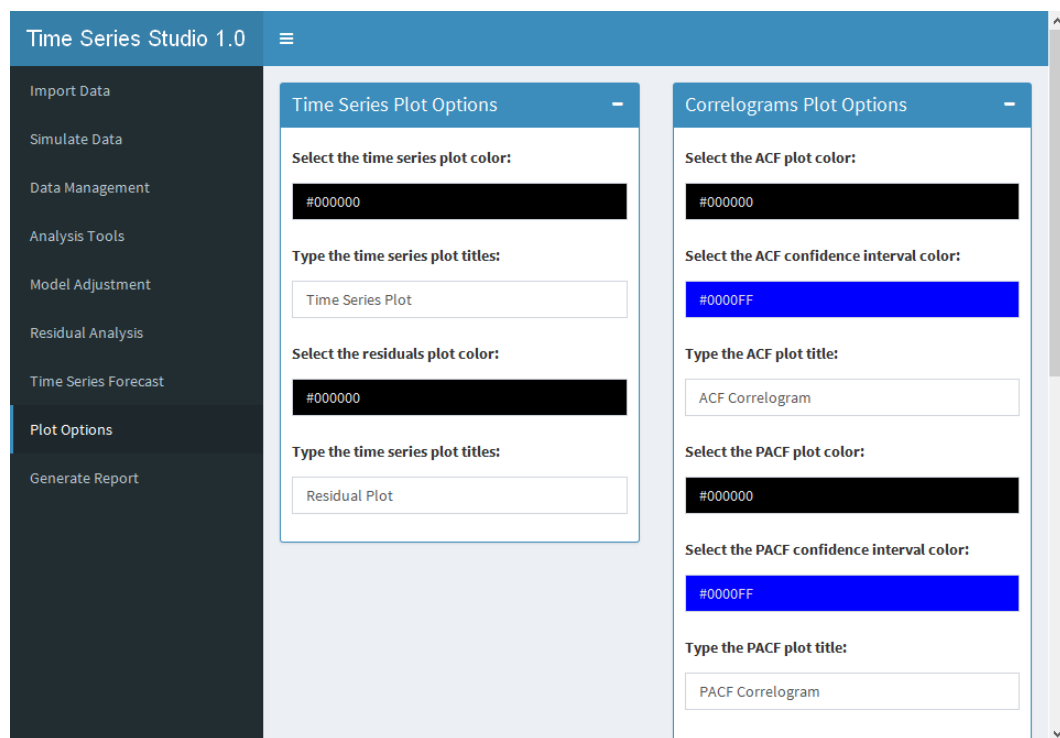


Figura 5.37: Tela de *Plot Options* com configurações de título e cor dos gráficos

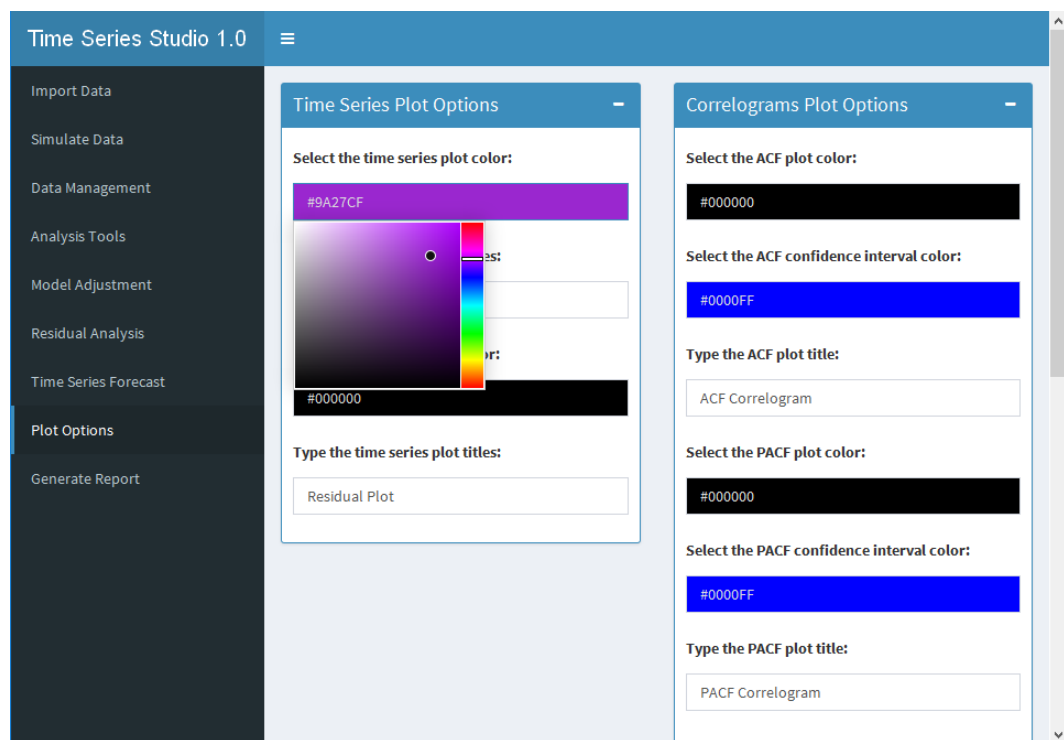


Figura 5.38: Tela de *Plot Options* com uma paleta de cor selecionada

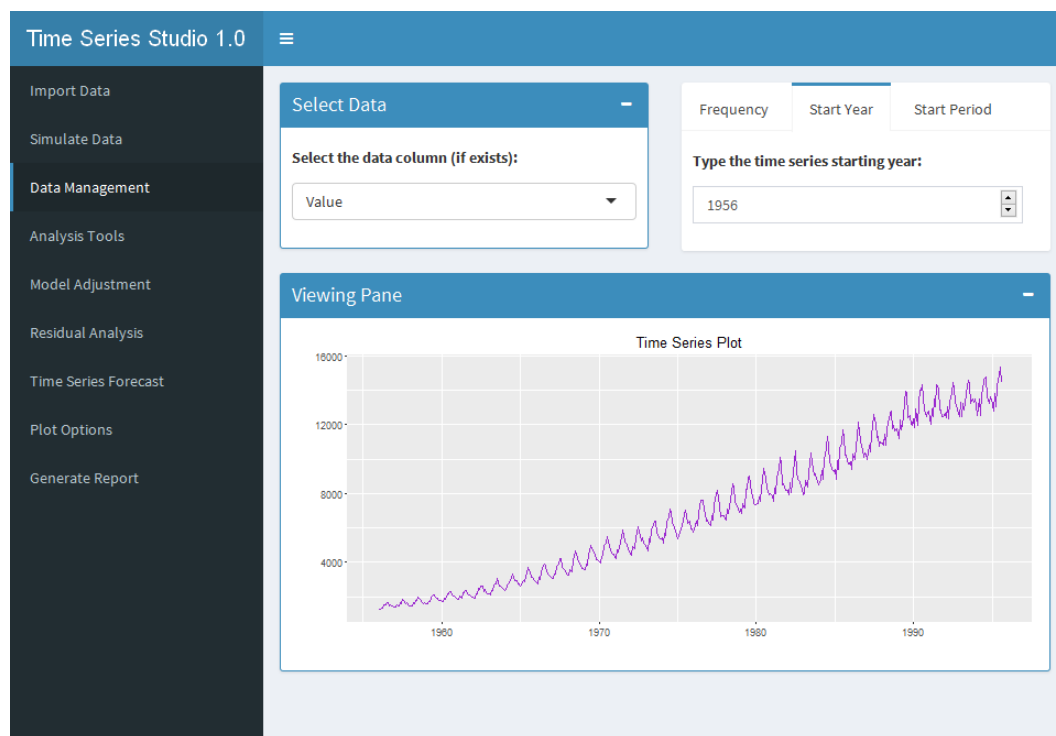


Figura 5.39: Tela de *Data Management* com a alteração da cor da série temporal

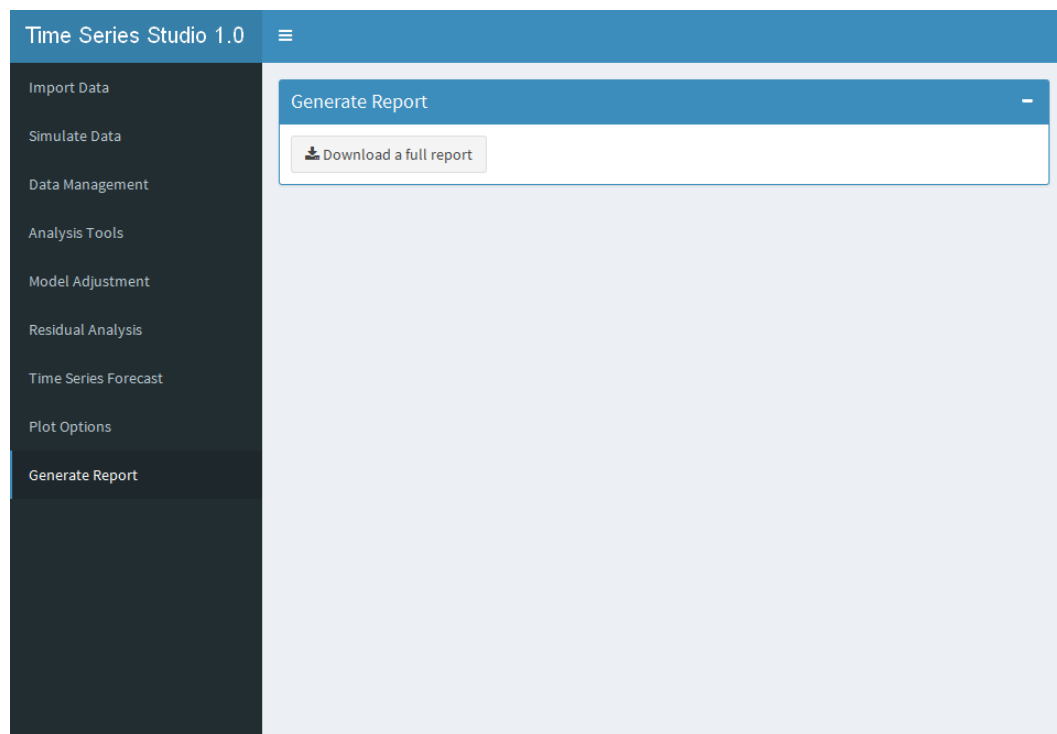


Figura 5.40: Tela de *Generate Report* com o botão para realização do download do relatório com os gráficos, resultados e saídas pertinentes

Capítulo 6

Referências Bibliográficas

SANTOS, P.T.G. dos; BREVI, F.P.; SILVA, J.F. da; COSTANZI FILHO, A.C.; FIALHO, D.M. **Desenvolvendo Interfaces Gráficas Utilizando Shiny**. SINAPE, 2016.

MONTGOMERY, D.C.; JENNINGS, C.L.; KULAHCI, M. **Time Series Analysis and Forecasting**. Wiley, 2ª Edição, 2015.

HYNDMAN, R.J.; KHANDAKHAR, Y. **Automatic Time Series Forecasting: The forecast Package of R**. Journal of Statistical Software, 2008, vol. 27, n. 3, p. 1–22.

MORANDAT, F; HILL, B.; OSVALD, L.; VITEK, J. **Evaluating the design of the R language: objects and functions for data analysis**. Purdue University, 2012.

IHAKA, R.; **R: Past and Future History**. Statistics Department, University of Auckland, 1998.

Home-page do **R-Project**. Disponível em: <https://www.r-project.org/about.html>. Acesso em 23/11/2016.

Teste de Phillips-Perron. Disponível em: <http://www.portalaction.com.br/series-temporais/142-teste-de-phillips-perron>. Acesso em 23/11/2016.

Teste de Dickey-Fuller Aumentado. Disponível em: <http://www.portalaction.com.br/series-temporais/141-teste-de-dickey-fuller-aumentado>. Acesso em 23/11/2016.

Teste de Box-Pierce. Disponível em: <http://www.portalaction.com.br/series-temporais/471-teste-de-box-pierce>. Acesso em 23/11/2016.

ARIMA modelling in R. Disponível em: <https://www.otexts.org/fpp/8/7>. Acesso em 24/11/2016.

Stationarity and differencing. Disponível em: <https://www.otexts.org/fpp/8/1>. Acesso em 24/11/2016.

Test Underlying Assumptions. Disponível em: <http://www.itl.nist.gov/div898/handbook/eda/section4/eda41.htm>. Acesso em 18/12/2016.

Introduction to ARIMA: nonseasonal models. Disponível em: <http://people.duke.edu/~rnau/411arim.htm>. Acesso em 25/11/2016.

BOLLERSLEV, T.; **Glossary to ARCH (GARCH).** Center for Research in Economic Analysis of Time Series (CREATES).

School of Economics and Management, University of Copenhagen, 2008.

KWIATKOWSKI, D; PHILLIPS, P.C.B.; SCHMIDT, P.; SHIN, Y. **Testing the null hypothesis of stationarity against the alternative of a unit root.**

Journal of Econometrics, 1992, vol. 54, p. 159–178. North-Holland, 1992.

HU, S. **Akaike Information Criterion.** Center for Research in Scientific Computation. North Carolina State University, 2007.